Creating Cubes for AVR On a Shoe-String Budget

3rd Annual PHIN Conference

Marion Kainer, M.D., M.P.H.

May 12, 2005

Outline

- Demonstration
- Definition of cube
- Shift in reporting paradigm
- Process steps to create a cube
- Gee whiz features
- Questions and Answers
- Contact details

Outline Page One

- Demo
  - Question 1
  - Question 2
  - Question 3
- Definitions

Outline Page Two

- The Reporting Paradigm Shift
- Process Steps
  - Step One
  - Step Two
  - Step Three
  - Step Four
  - Step Five
  - Step Six
  - Step Seven
- Gee Whiz Features
- Q & A

Demo

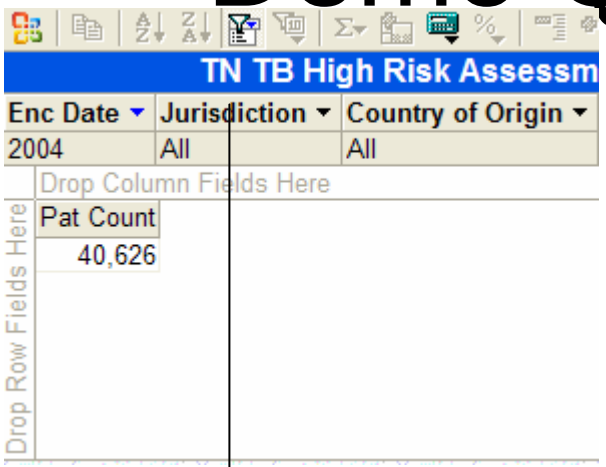
- Data set used
  - Tennessee's Tuberculosis Risk Assessment Tool (The RAT)
    - Calendar year 2004
    - 40,626 risk assessments
    - 70 variables per assessment

Question One

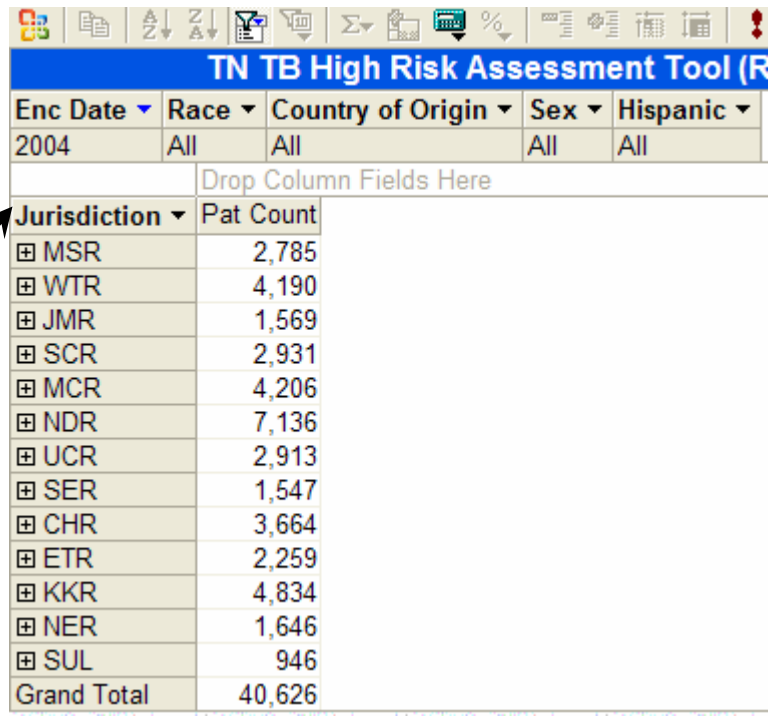
- Are our field staff using the results of the RAT to perform tuberculin skin tests appropriately?

  i.e.,
- Test those assessed at high risk of TB infection
- Not test those assessed at low risk of TB infection

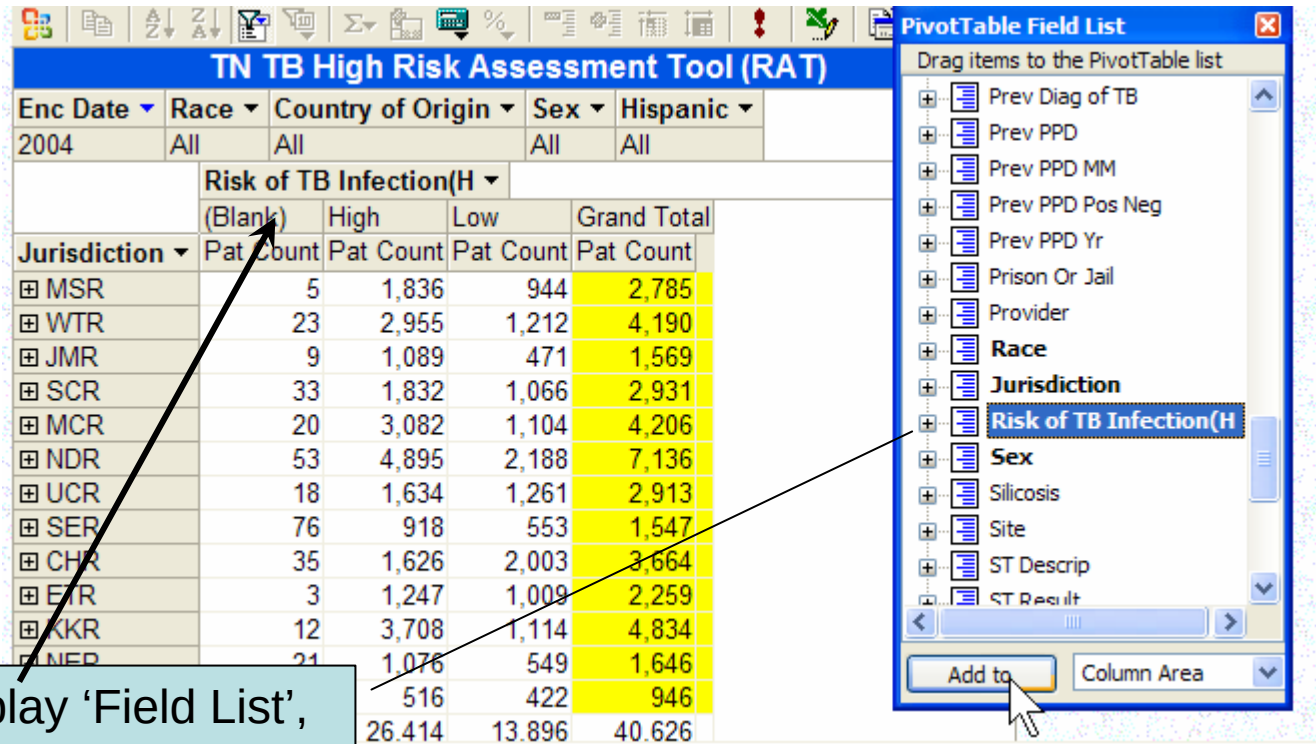
Demo Question One



1. Drag and drop Jurisdiction from Filter to Row Area

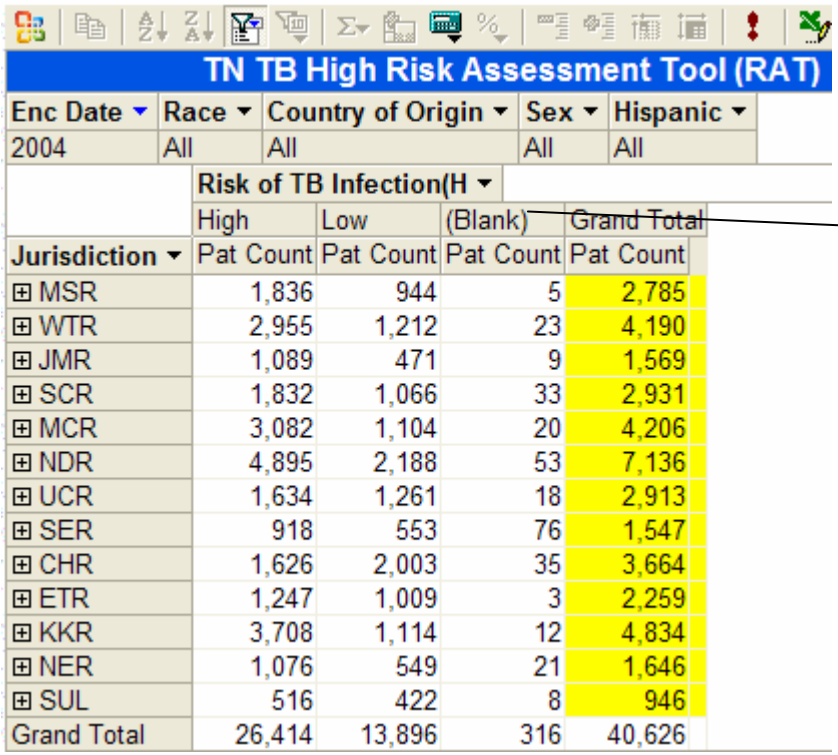


Demo Question One



2. Display 'Field List', highlight '**Risk of TB Infection**' field and drag and drop to column area

Demo Question One



TN TB High Risk Assessment Tool (RAT)

Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

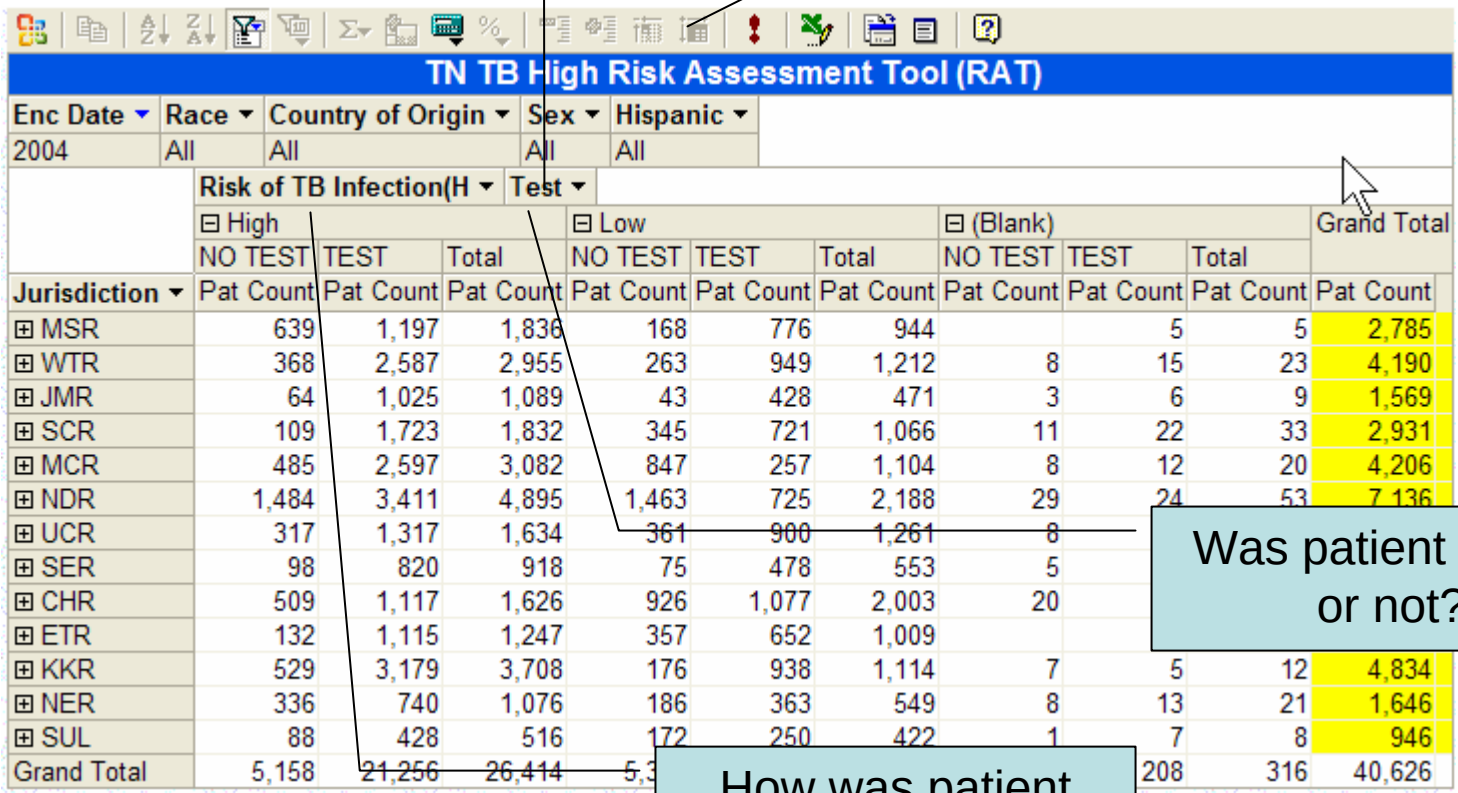
	Risk of TB Infection(H			
	High	Low	(Blank)	Grand Total
Jurisdiction	Pat Count	Pat Count	Pat Count	Pat Count
⊞ MSR	1,836	944	5	2,785
⊞ WTR	2,955	1,212	23	4,190
⊞ JMR	1,089	471	9	1,569
⊞ SCR	1,832	1,066	33	2,931
⊞ MCR	3,082	1,104	20	4,206
⊞ NDR	4,895	2,188	53	7,136
⊞ UCR	1,634	1,261	18	2,913
⊞ SER	918	553	76	1,547
⊞ CHR	1,626	2,003	35	3,664
⊞ ETR	1,247	1,009	3	2,259
⊞ KKR	3,708	1,114	12	4,834
⊞ NER	1,076	549	21	1,646
⊞ SUL	516	422	8	946
Grand Total	26,414	13,896	316	40,626

3. Rearrange columns by dragging and dropping

o Question One

4. Insert the variable 'Test' into column area from Field List

5. Drill into column details



TN TB High Risk Assessment Tool (RAT)

Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

	Risk of TB Infection(H	Test								
	⊟ High			⊟ Low			⊟ (Blank)			Grand Total
	NO TEST	TEST	Total	NO TEST	TEST	Total	NO TEST	TEST	Total	
Jurisdiction	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
⊞ MSR	639	1,197	1,836	168	776	944		5	5	2,785
⊞ WTR	368	2,587	2,955	263	949	1,212	8	15	23	4,190
⊞ JMR	64	1,025	1,089	43	428	471	3	6	9	1,569
⊞ SCR	109	1,723	1,832	345	721	1,066	11	22	33	2,931
⊞ MCR	485	2,597	3,082	847	257	1,104	8	12	20	4,206
⊞ NDR	1,484	3,411	4,895	1,463	725	2,188	29	24	53	7,136
⊞ UCR	317	1,317	1,634	361	900	1,261	8			
⊞ SER	98	820	918	75	478	553	5			
⊞ CHR	509	1,117	1,626	926	1,077	2,003	20			
⊞ ETR	132	1,115	1,247	357	652	1,009				
⊞ KKR	529	3,179	3,708	176	938	1,114	7	5	12	4,834
⊞ NER	336	740	1,076	186	363	549	8	13	21	1,646
⊞ SUL	88	428	516	172	250	422	1	7	8	946
Grand Total	5,158	21,256	26,414	5,3			208		316	40,626

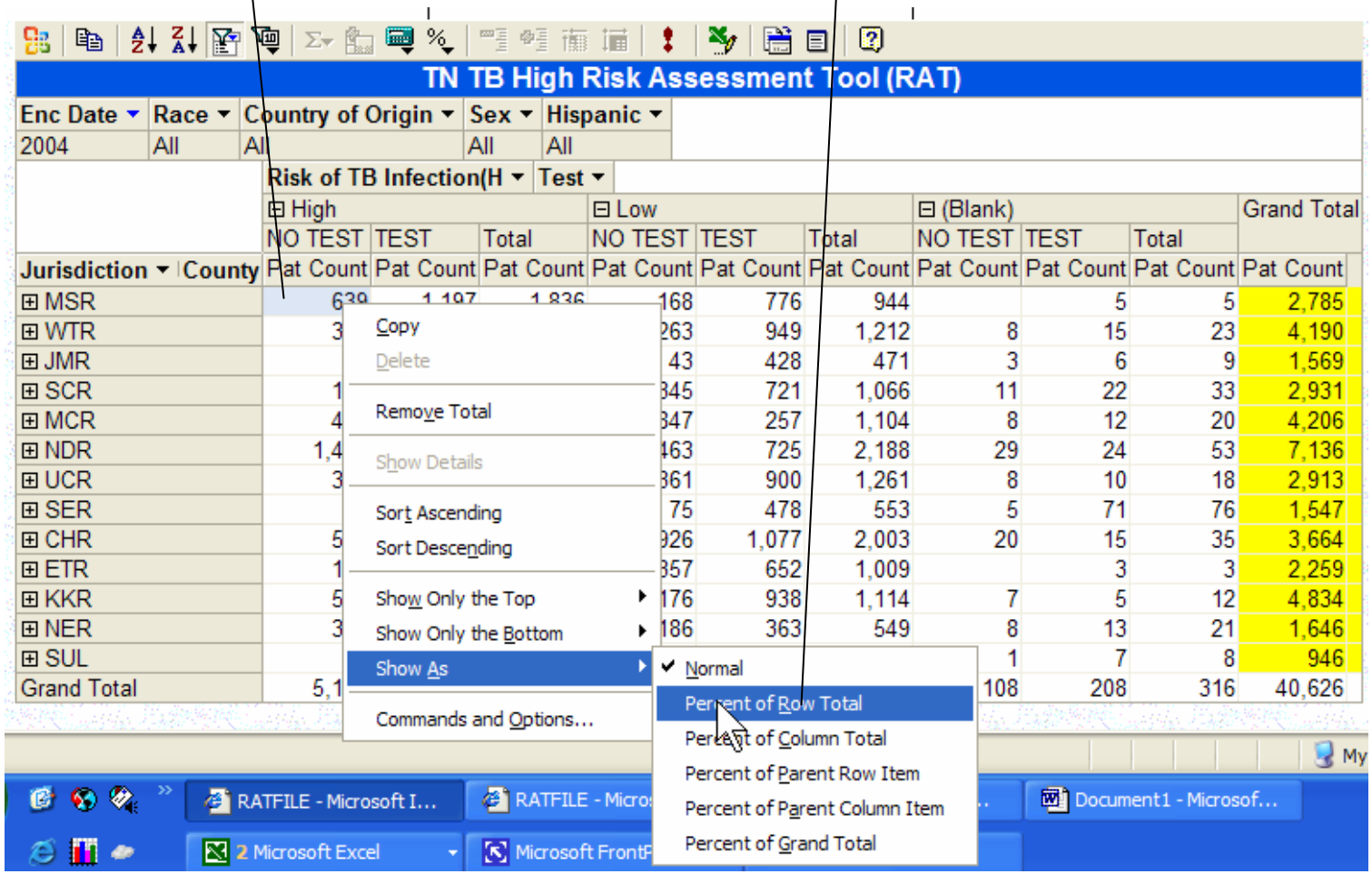
Was patient tested or not?

How was patient assessed by the RAT?

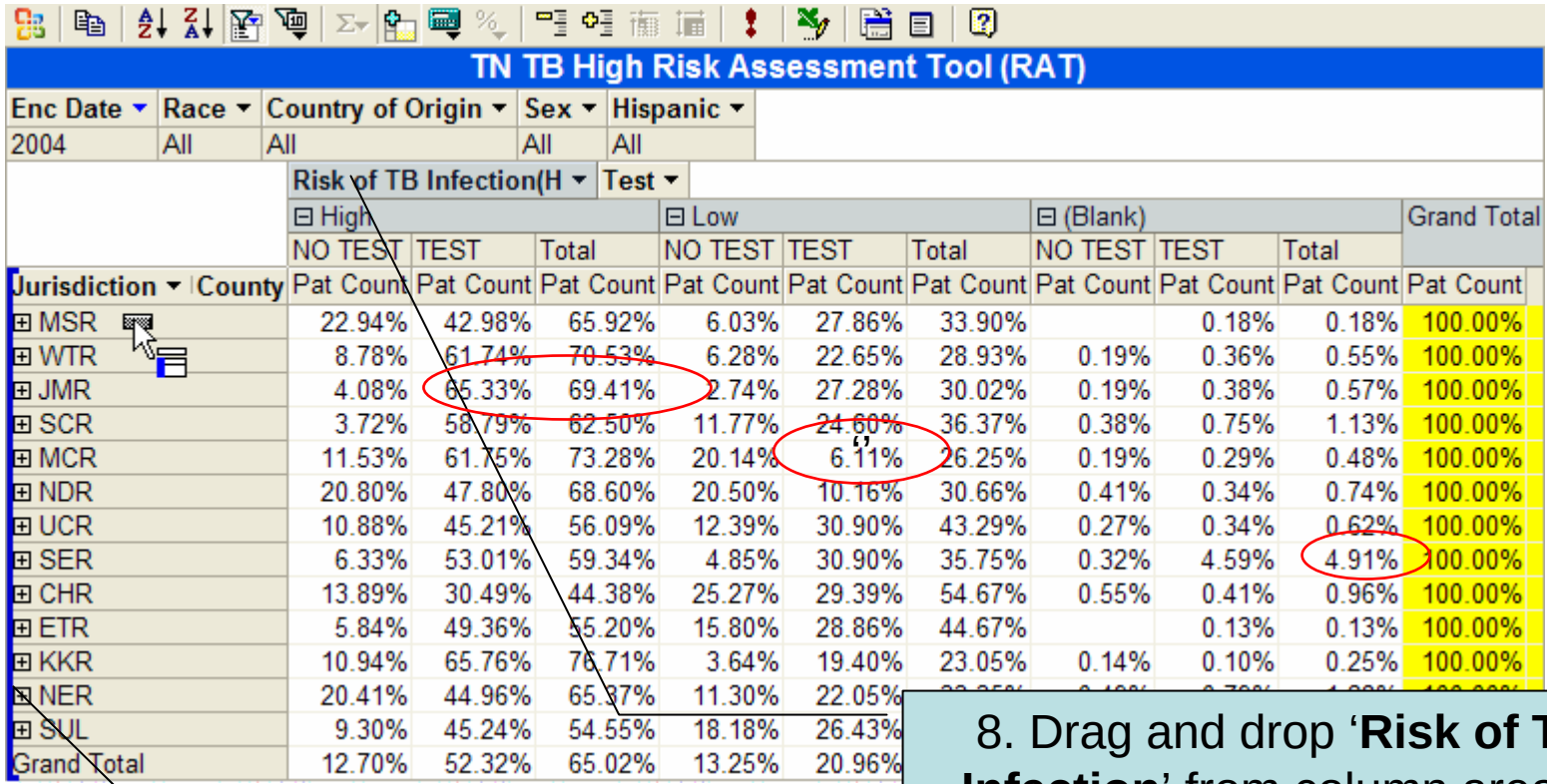
6. Right click on any data cell

mo Question

7. Select 'Show as→ Percent of Row Total'



Demo Question One



TN TB High Risk Assessment Tool (RAT)

Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

	Risk of TB Infection(H	High			Low			(Blank)			Grand Total
	Test	NO TEST	TEST	Total	NO TEST	TEST	Total	NO TEST	TEST	Total	
Jurisdiction	County	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
MSR		22.94%	42.98%	65.92%	6.03%	27.86%	33.90%		0.18%	0.18%	100.00%
WTR		8.78%	61.74%	70.53%	6.28%	22.65%	28.93%	0.19%	0.36%	0.55%	100.00%
JMR		4.08%	65.33%	69.41%	2.74%	27.28%	30.02%	0.19%	0.38%	0.57%	100.00%
SCR		3.72%	58.79%	62.50%	11.77%	24.60%	36.37%	0.38%	0.75%	1.13%	100.00%
MCR		11.53%	61.75%	73.28%	20.14%	6.11%	26.25%	0.19%	0.29%	0.48%	100.00%
NDR		20.80%	47.80%	68.60%	20.50%	10.16%	30.66%	0.41%	0.34%	0.74%	100.00%
UCR		10.88%	45.21%	56.09%	12.39%	30.90%	43.29%	0.27%	0.34%	0.62%	100.00%
SER		6.33%	53.01%	59.34%	4.85%	30.90%	35.75%	0.32%	4.59%	4.91%	100.00%
CHR		13.89%	30.49%	44.38%	25.27%	29.39%	54.67%	0.55%	0.41%	0.96%	100.00%
ETR		5.84%	49.36%	55.20%	15.80%	28.86%	44.67%		0.13%	0.13%	100.00%
KKR		10.94%	65.76%	76.71%	3.64%	19.40%	23.05%	0.14%	0.10%	0.25%	100.00%
NER		20.41%	44.96%	65.37%	11.30%	22.05%					
SUL		9.30%	45.24%	54.55%	18.18%	26.43%					
Grand Total		12.70%	52.32%	65.02%	13.25%	20.96%					

8. Drag and drop '**Risk of TB Infection**' from column area to Row Area

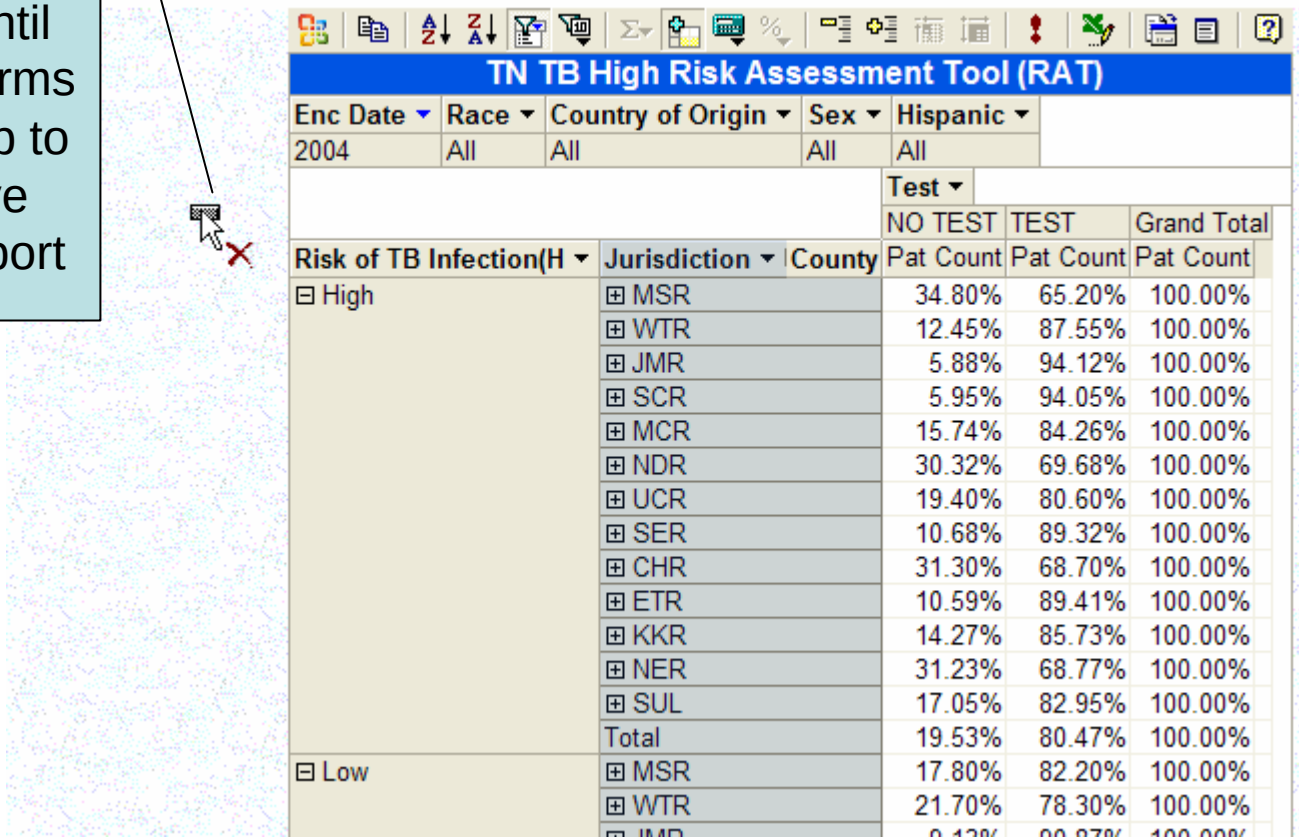
9.[indicates where the field will drop

Question Two

- How well does the use of Tennessee's RAT predict a positive Tuberculin Skin Test outcome?

Demo Question Two

10. Drag Jurisdiction outside table until an 'X' forms and drop to remove from report

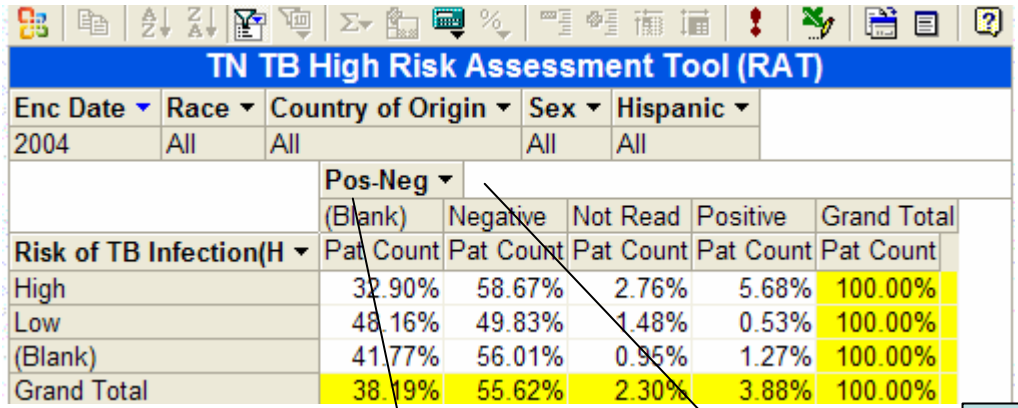


TN TB High Risk Assessment Tool (RAT)

Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

			Test		
			NO TEST	TEST	Grand Total
Risk of TB Infection(H	Jurisdiction	County	Pat Count	Pat Count	Pat Count
⊟ High	⊞ MSR		34.80%	65.20%	100.00%
	⊞ WTR		12.45%	87.55%	100.00%
	⊞ JMR		5.88%	94.12%	100.00%
	⊞ SCR		5.95%	94.05%	100.00%
	⊞ MCR		15.74%	84.26%	100.00%
	⊞ NDR		30.32%	69.68%	100.00%
	⊞ UCR		19.40%	80.60%	100.00%
	⊞ SER		10.68%	89.32%	100.00%
	⊞ CHR		31.30%	68.70%	100.00%
	⊞ ETR		10.59%	89.41%	100.00%
	⊞ KKR		14.27%	85.73%	100.00%
	⊞ NER		31.23%	68.77%	100.00%
	⊞ SUL		17.05%	82.95%	100.00%
	Total		19.53%	80.47%	100.00%
⊟ Low	⊞ MSR		17.80%	82.20%	100.00%
	⊞ WTR		21.70%	78.30%	100.00%

Demo Question Two



TN TB High Risk Assessment Tool (RAT)

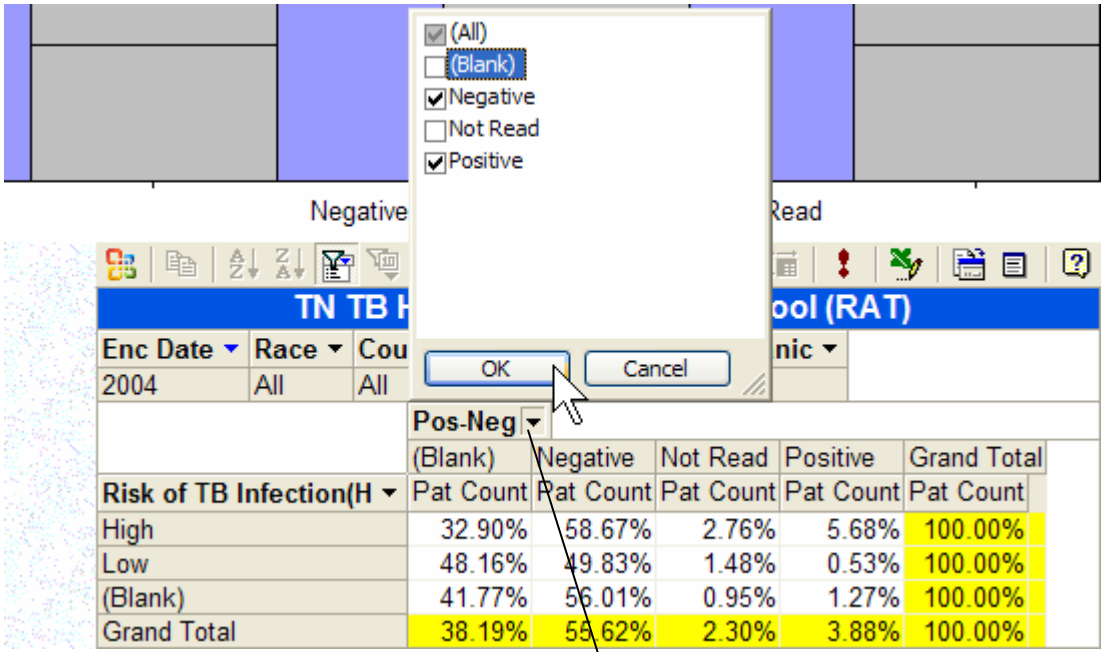
Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

	Pos-Neg				
	(Blank)	Negative	Not Read	Positive	Grand Total
Risk of TB Infection(H	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
High	32.90%	58.67%	2.76%	5.68%	100.00%
Low	48.16%	49.83%	1.48%	0.53%	100.00%
(Blank)	41.77%	56.01%	0.95%	1.27%	100.00%
Grand Total	38.19%	55.62%	2.30%	3.88%	100.00%

11. Remove the field '**Test**' and add field '**Pos-Neg**' to the column area

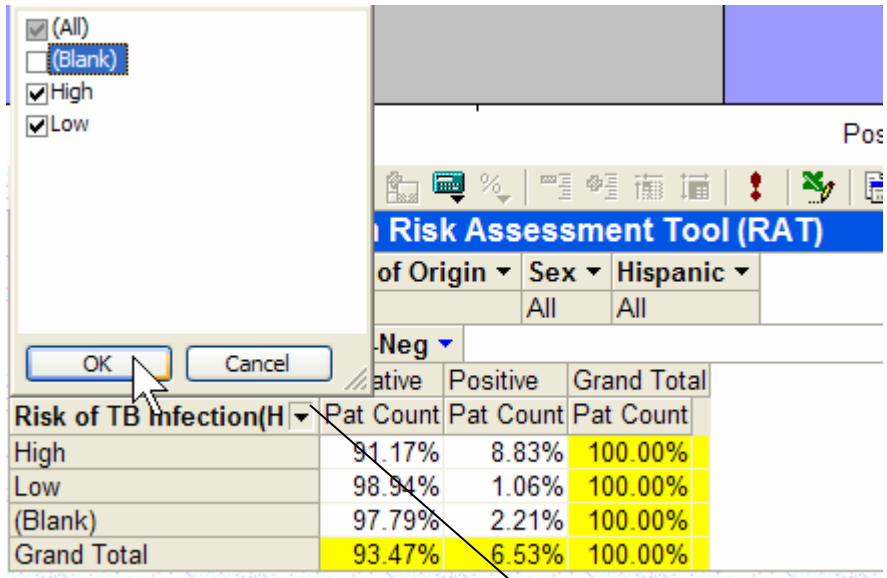
'Pos-Neg' indicates the result of a skin test.

Demo Question Two



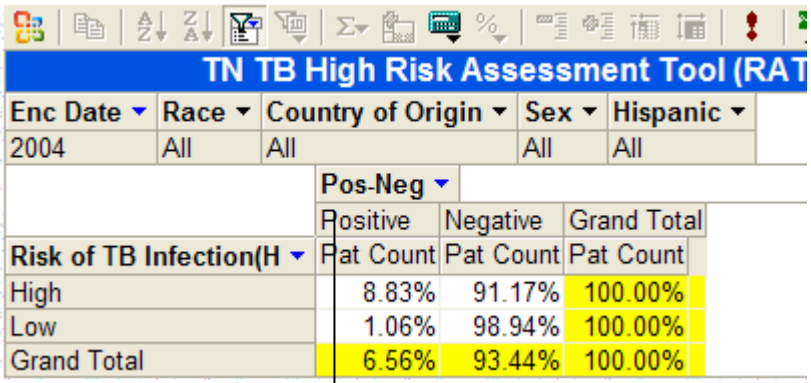
12. Click the down arrow by ‘Pos-Neg’ and deselect column values (Blank) and ‘Not Read’ to omit from report

Demo Question Two



13. Click the down arrow by ‘Risk of TB’ and deselect row values to omit from report

Demo Question Two



TN TB High Risk Assessment Tool (RAT

Enc Date	Race	Country of Origin	Sex	Hispanic
2004	All	All	All	All

	Pos-Neg		
	Positive	Negative	Grand Total
Risk of TB Infection(H	Pat Count	Pat Count	Pat Count
High	8.83%	91.17%	100.00%
Low	1.06%	98.94%	100.00%
Grand Total	6.56%	93.44%	100.00%

14. Drag and drop data value to rearrange columns

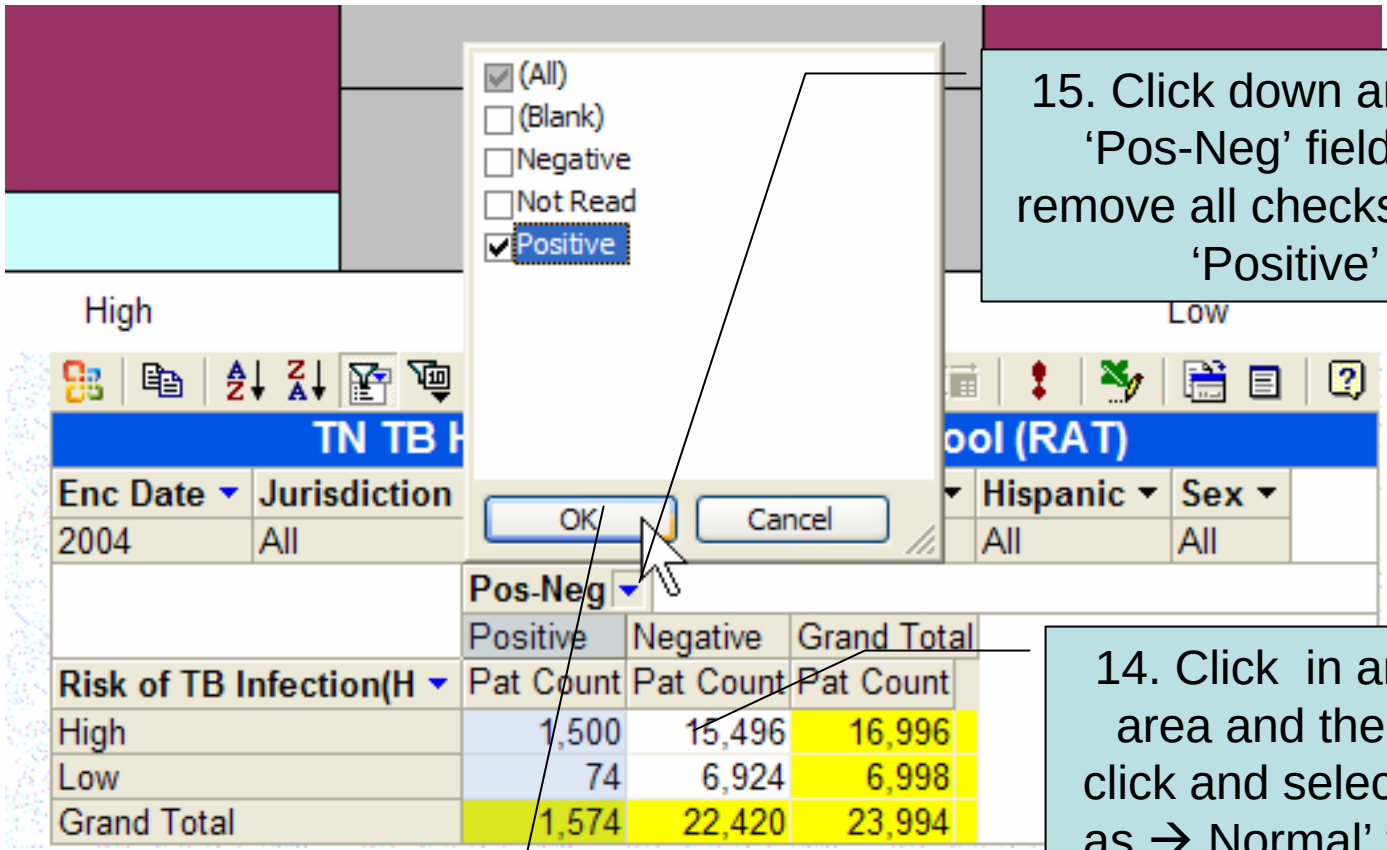
Question Three

- How do the non-USA born High Risk and Positive skin tests break down by country, ethnicity, and gender?

or

- What country, ethnicity, and gender contribute the most positive TST high risk cases?

Demo Question Three

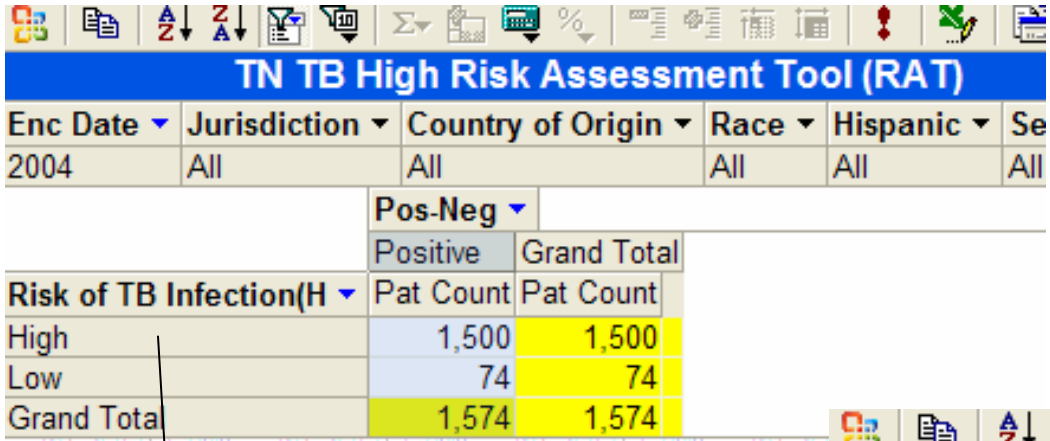


15. Click down arrow by 'Pos-Neg' field and remove all checks except 'Positive'

14. Click in any data area and then right click and select 'Show as → Normal' to show actual numbers

16. Click 'OK' to continue

Demo Question Three



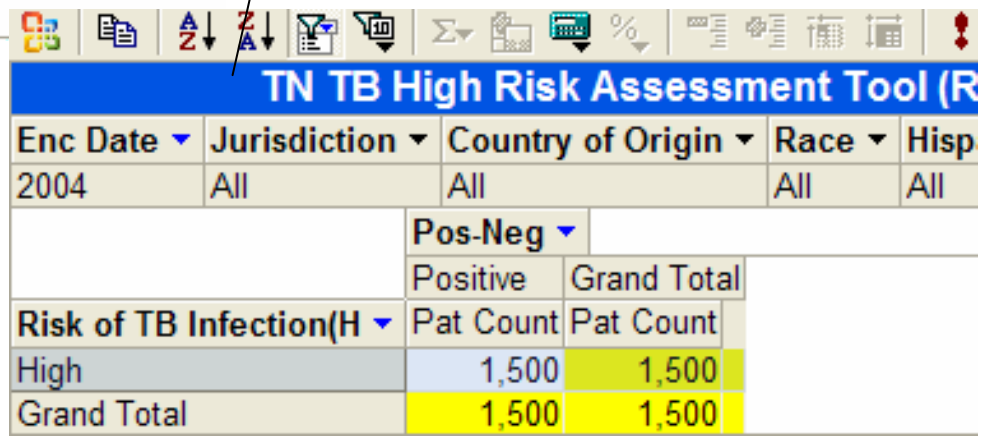
TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Country of Origin	Race	Hispanic	Se
2004	All	All	All	All	All

	Pos-Neg	
	Positive	Grand Total
Risk of TB Infection(H	Pat Count	Pat Count
High	1,500	1,500
Low	74	74
Grand Total	1,574	1,574

17. Repeat steps 15 and 16 and select the 'High Risk' row in 'Risk of TB Infection'

18. Ok now we are going to investigate these 1,500 individuals

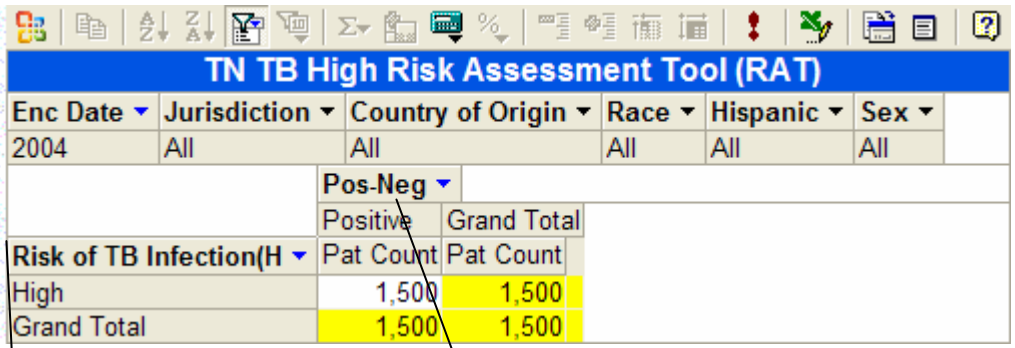


TN TB High Risk Assessment Tool (R

Enc Date	Jurisdiction	Country of Origin	Race	Hisp
2004	All	All	All	All

	Pos-Neg	
	Positive	Grand Total
Risk of TB Infection(H	Pat Count	Pat Count
High	1,500	1,500
Grand Total	1,500	1,500

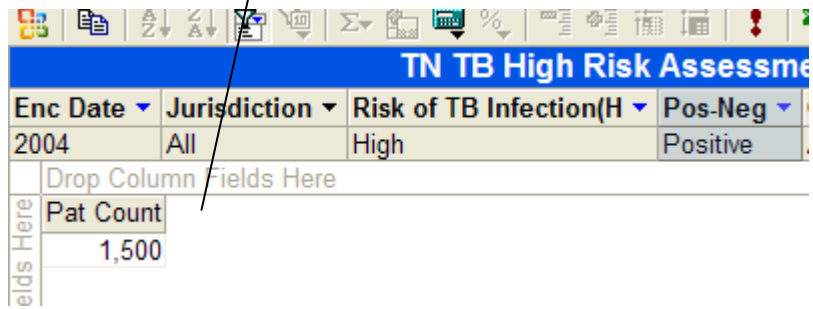
Demo Question Three



TN TB High Risk Assessment Tool (RAT)					
Enc Date	Jurisdiction	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All	All

	Pos-Neg	
	Positive	Grand Total
Risk of TB Infection(H	Pat Count	Pat Count
High	1,500	1,500
Grand Total	1,500	1,500

20. Our new population



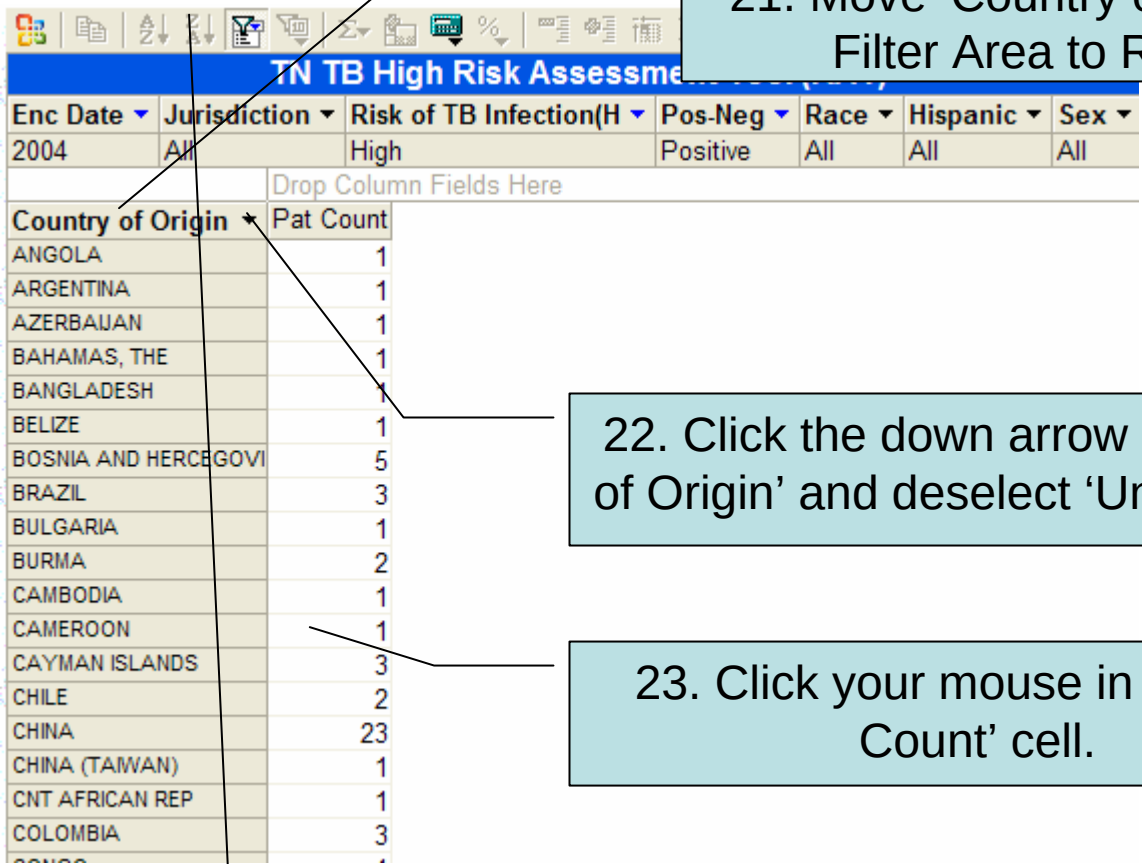
TN TB High Risk Assessme			
Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg
2004	All	High	Positive

Drop Column Fields Here

elds Here	Pat Count
	1,500

19. Now move both ‘Risk of TB Infection’ and ‘Pos-Neg’ to Filter Area

Demo Question Three



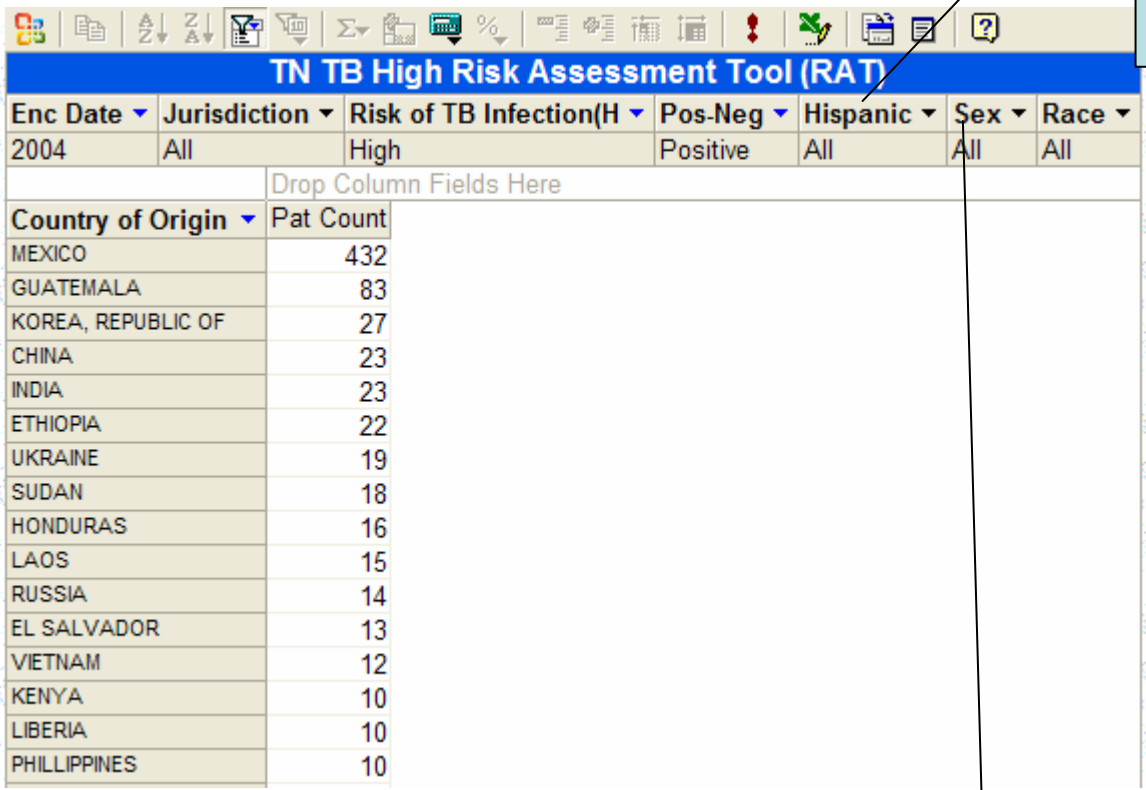
21. Move ‘Country of Origin’ from Filter Area to Row Area

22. Click the down arrow by ‘Country of Origin’ and deselect ‘United States’

23. Click your mouse in any ‘Pat Count’ cell.

24. Click the ‘Sort Descending button on Toolbar.

Demo Question Three



TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg	Hispanic	Sex	Race
2004	All	High	Positive	All	All	All

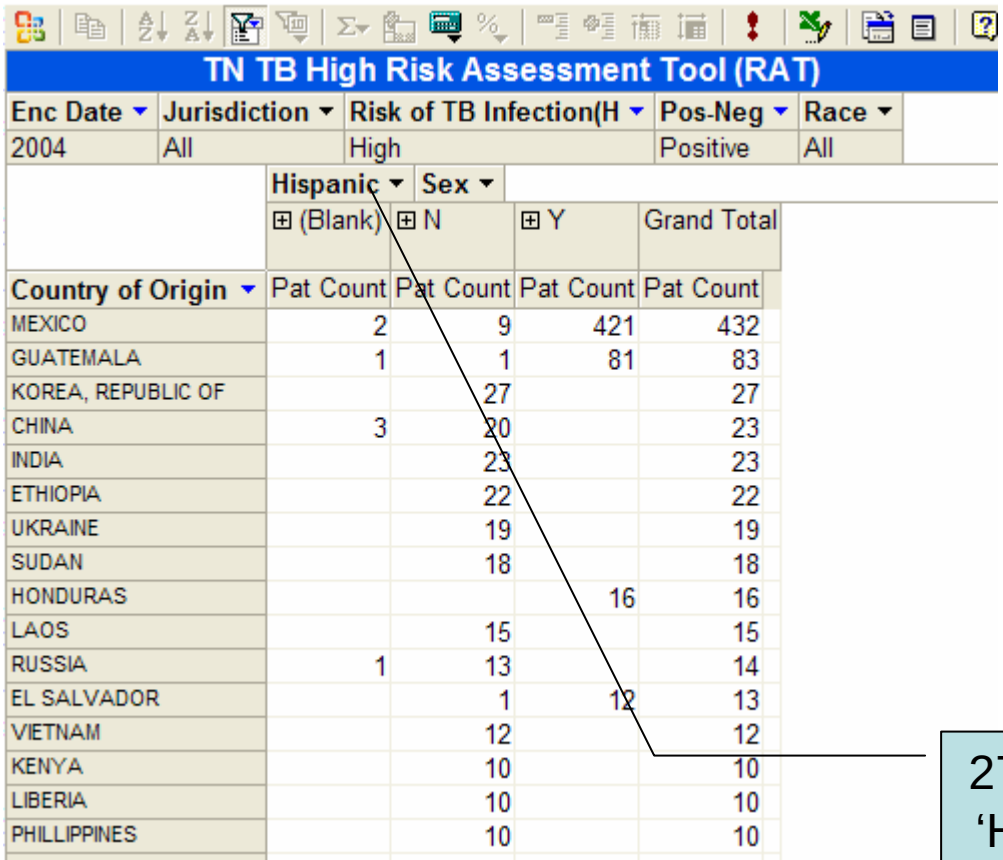
Drop Column Fields Here

Country of Origin	Pat Count
MEXICO	432
GUATEMALA	83
KOREA, REPUBLIC OF	27
CHINA	23
INDIA	23
ETHIOPIA	22
UKRAINE	19
SUDAN	18
HONDURAS	16
LAOS	15
RUSSIA	14
EL SALVADOR	13
VIETNAM	12
KENYA	10
LIBERIA	10
PHILLIPPINES	10

25. Move ‘Hispanic from Filter Area to Column Area

26. Move ‘Sex from Filter Area to Column Area to the right of ‘Hispanic’

Demo Question Three



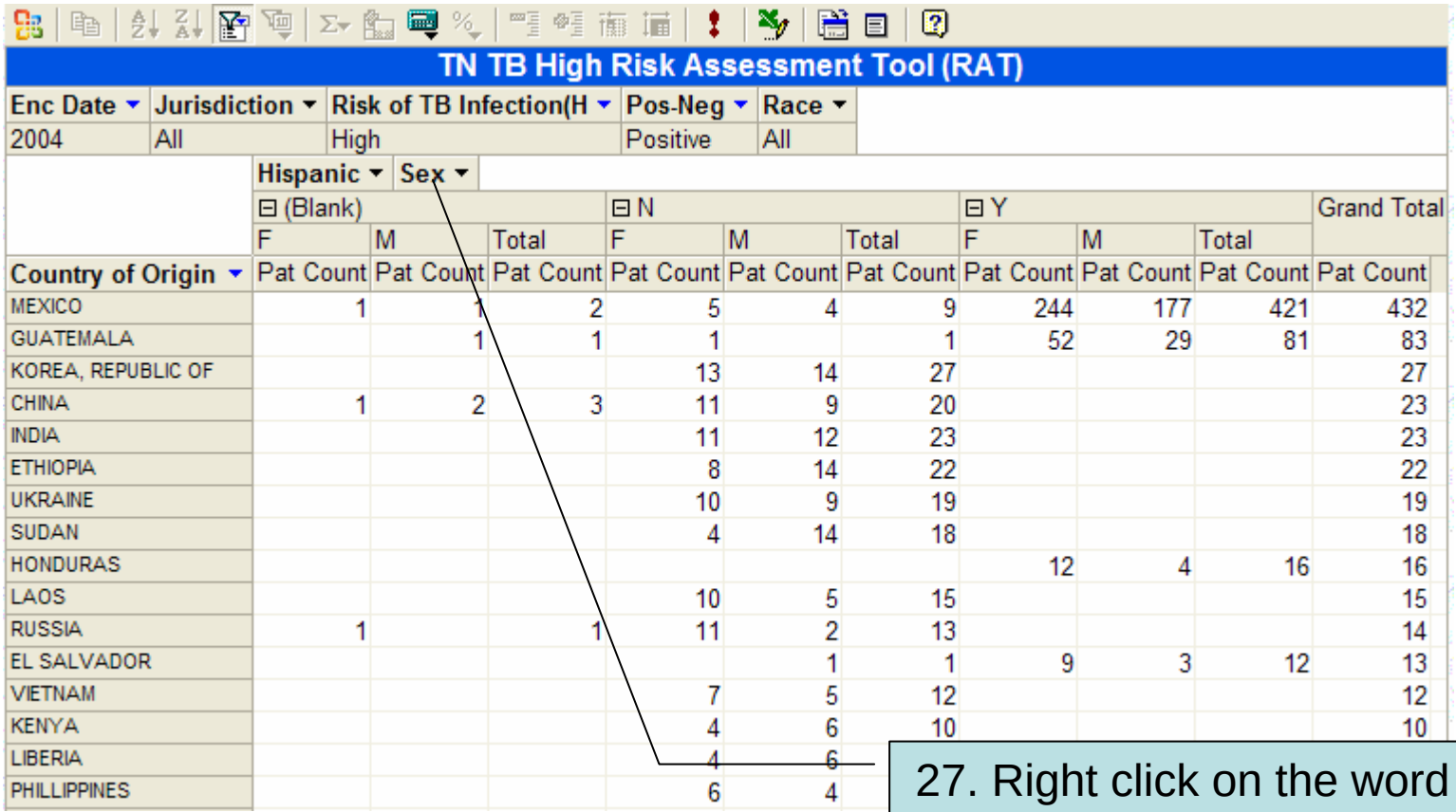
TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg	Race
2004	All	High	Positive	All

	Hispanic: ⊞ (Blank)	Hispanic: ⊞ N	Hispanic: ⊞ Y	Grand Total
Country of Origin	Pat Count	Pat Count	Pat Count	Pat Count
MEXICO	2	9	421	432
GUATEMALA	1	1	81	83
KOREA, REPUBLIC OF		27		27
CHINA	3	20		23
INDIA		23		23
ETHIOPIA		22		22
UKRAINE		19		19
SUDAN		18		18
HONDURAS			16	16
LAOS		15		15
RUSSIA	1	13		14
EL SALVADOR		1	12	13
VIETNAM		12		12
KENYA		10		10
LIBERIA		10		10
PHILLIPPINES		10		10

27. Right click on the word ‘Hispanic’ in Column Area and select ‘Expand Items’

Demo Question Three



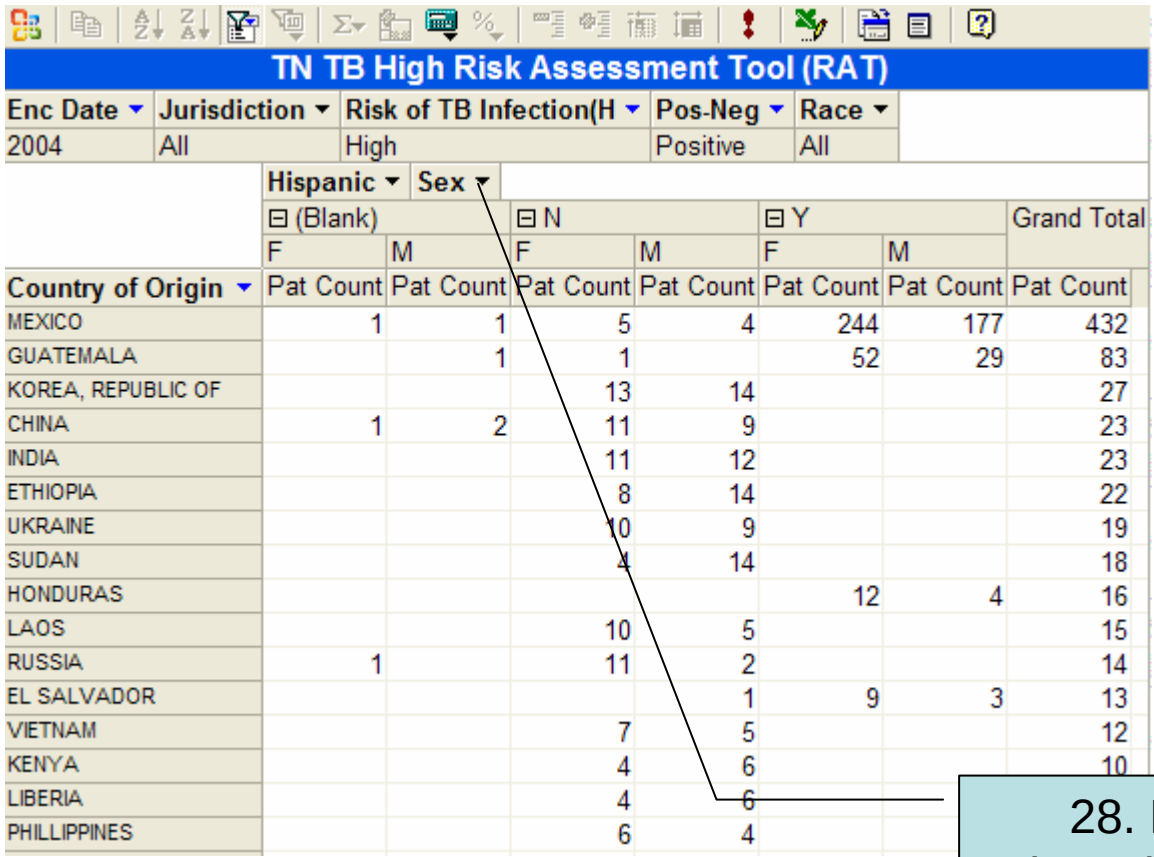
TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg	Race
2004	All	High	Positive	All

	Hispanic	(Blank)			N			Y			Grand Total
	Sex	F	M	Total	F	M	Total	F	M	Total	
Country of Origin		Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
MEXICO		1	1	2	5	4	9	244	177	421	432
GUATEMALA			1	1	1		1	52	29	81	83
KOREA, REPUBLIC OF					13	14	27				27
CHINA		1	2	3	11	9	20				23
INDIA					11	12	23				23
ETHIOPIA					8	14	22				22
UKRAINE					10	9	19				19
SUDAN					4	14	18				18
HONDURAS								12	4	16	16
LAOS					10	5	15				15
RUSSIA		1		1	11	2	13				14
EL SALVADOR						1	1	9	3	12	13
VIETNAM					7	5	12				12
KENYA					4	6	10				10
LIBERIA					4	6					
PHILLIPPINES					6	4					

27. Right click on the word ‘Sex’ in Column Area and uncheck ‘Subtotal’

Demo Question Three



TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg	Race
2004	All	High	Positive	All

	Hispanic / Sex	(Blank)		N		Y		Grand Total
		F	M	F	M	F	M	
Country of Origin		Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
MEXICO		1	1	5	4	244	177	432
GUATEMALA			1	1		52	29	83
KOREA, REPUBLIC OF				13	14			27
CHINA		1	2	11	9			23
INDIA				11	12			23
ETHIOPIA				8	14			22
UKRAINE				10	9			19
SUDAN				4	14			18
HONDURAS						12	4	16
LAOS				10	5			15
RUSSIA		1		11	2			14
EL SALVADOR					1	9	3	13
VIETNAM				7	5			12
KENYA				4	6			10
LIBERIA				4	6			
PHILLIPPINES				6	4			

28. Now rearrange Hispanic Columns to ‘Y’ – ‘N’ and ‘(blank)’

Demo Question Three

TN TB High Risk Assessment Tool (RAT)

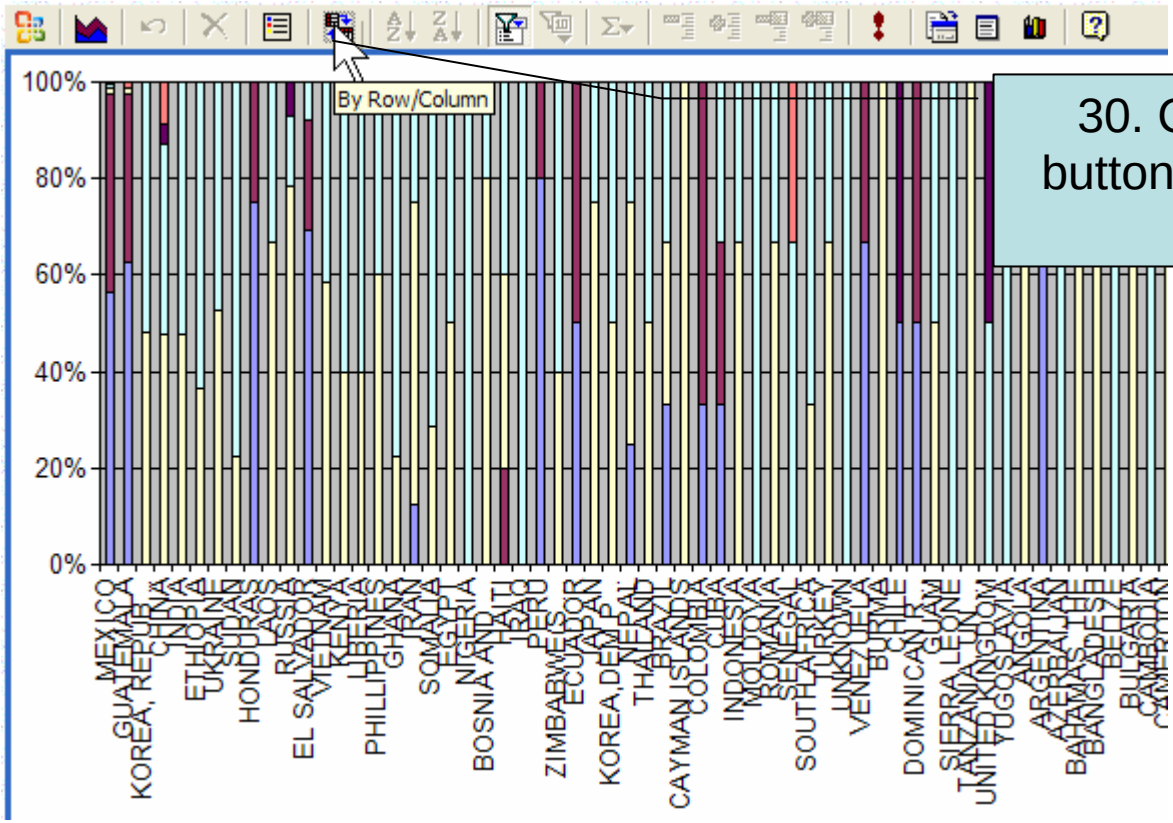
Enc Date	Jurisdiction	Risk of TB Infection(H	Pos-Neg	Race
2004	All	High	Positive	All

	Hispanic						
	⊟ Y		⊟ N		⊟ (Blank)		Grand Total
	F	M	F	M	F	M	
Country of Origin	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count	Pat Count
MEXICO	244	177	5	4	1	1	432
GUATEMALA	52	29	1			1	83
KOREA, REPUBLIC OF			13	14			27
CHINA			11	9	1	2	23
INDIA			11	12			23
ETHIOPIA			8	14			22
UKRAINE			10	9			19
SUDAN			4	14			18
HONDURAS	12	4					16
LAOS			10	5			15
RUSSIA			11	2	1		14
EL SALVADOR	9	3		1			13
VIETNAM			7	5			12
KENYA			4	6			10
LIBERIA			4	6			10
PHILLIPPINES			6	4			10

28. TAH DAH !!!

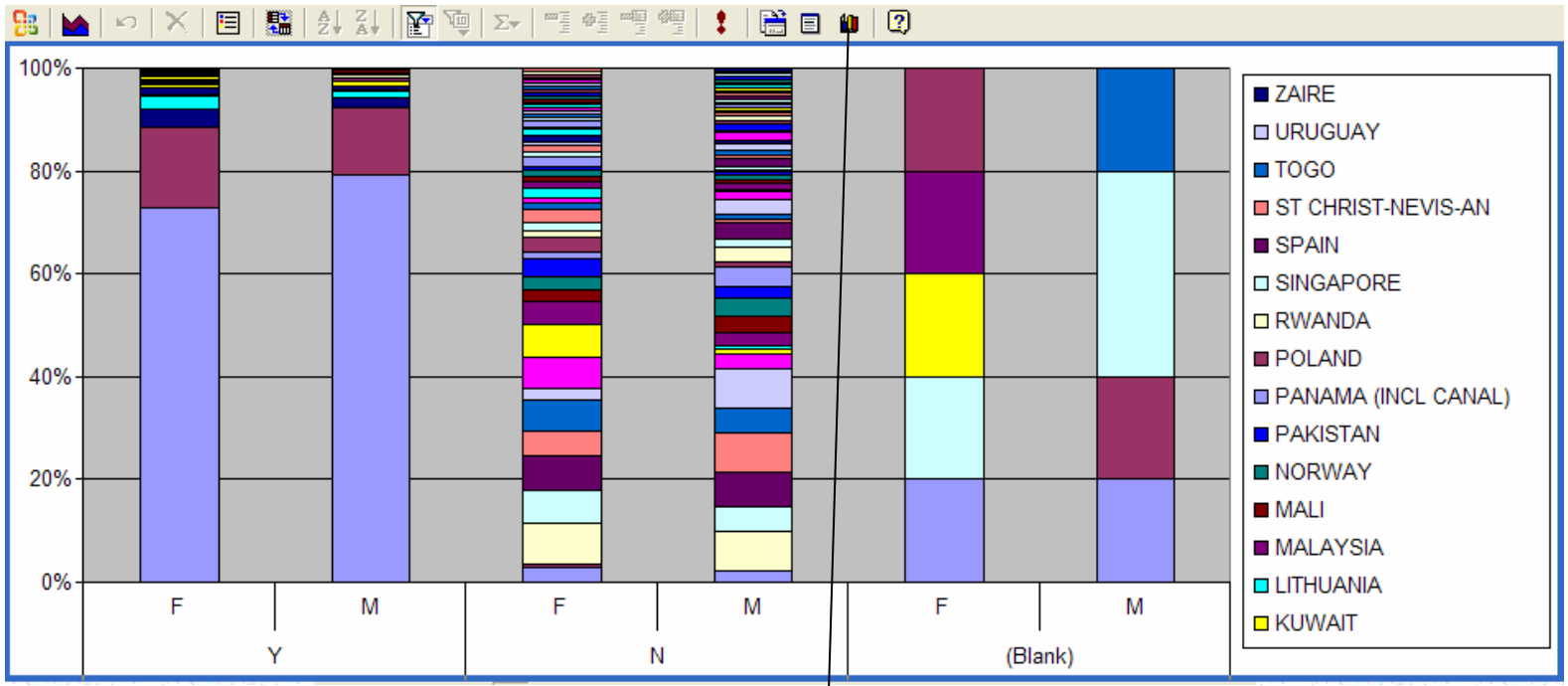
Demo Questior

29. Now let us make the chart tell us something



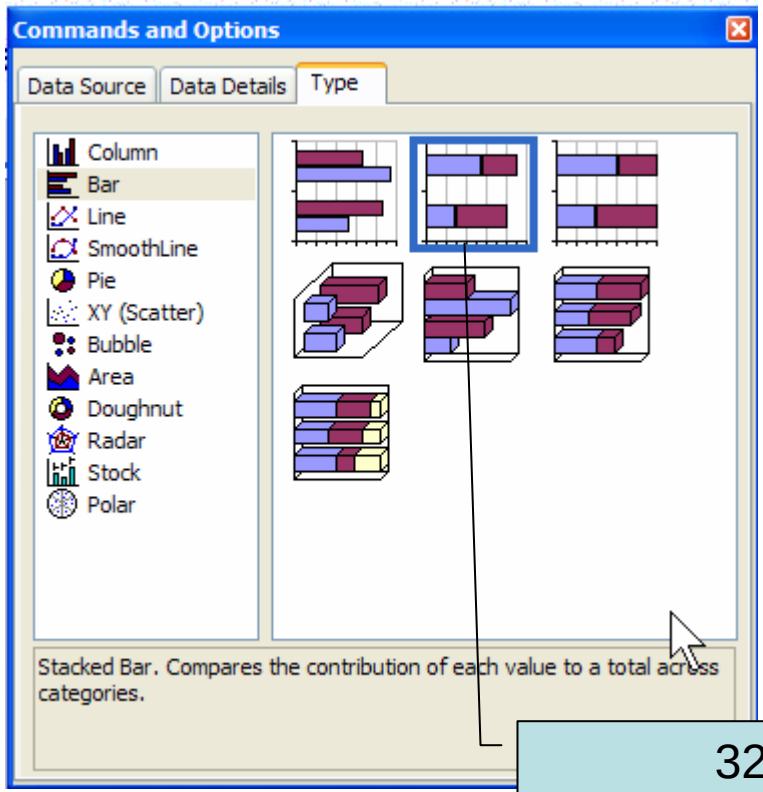
30. Click the ‘By Row/Column button on chart toolbar to change direction of chart

Demo Question Three



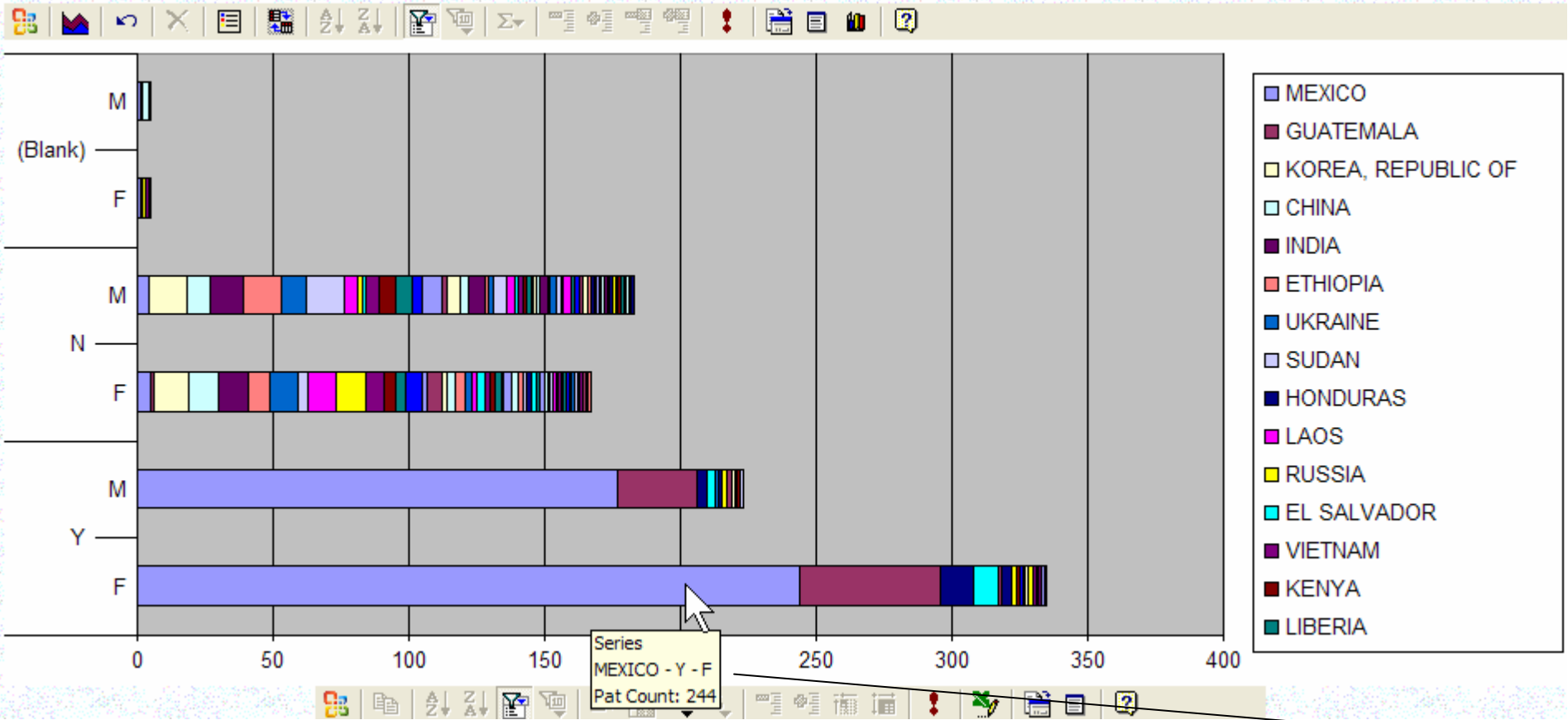
31. Click the Chart Wizard button.

Demo Question Three



32. Click the ‘Type’ tab
33. Select ‘Bar’ from left window
33. Click the ‘Stacked Bar’ in right window
33. Close wizard window

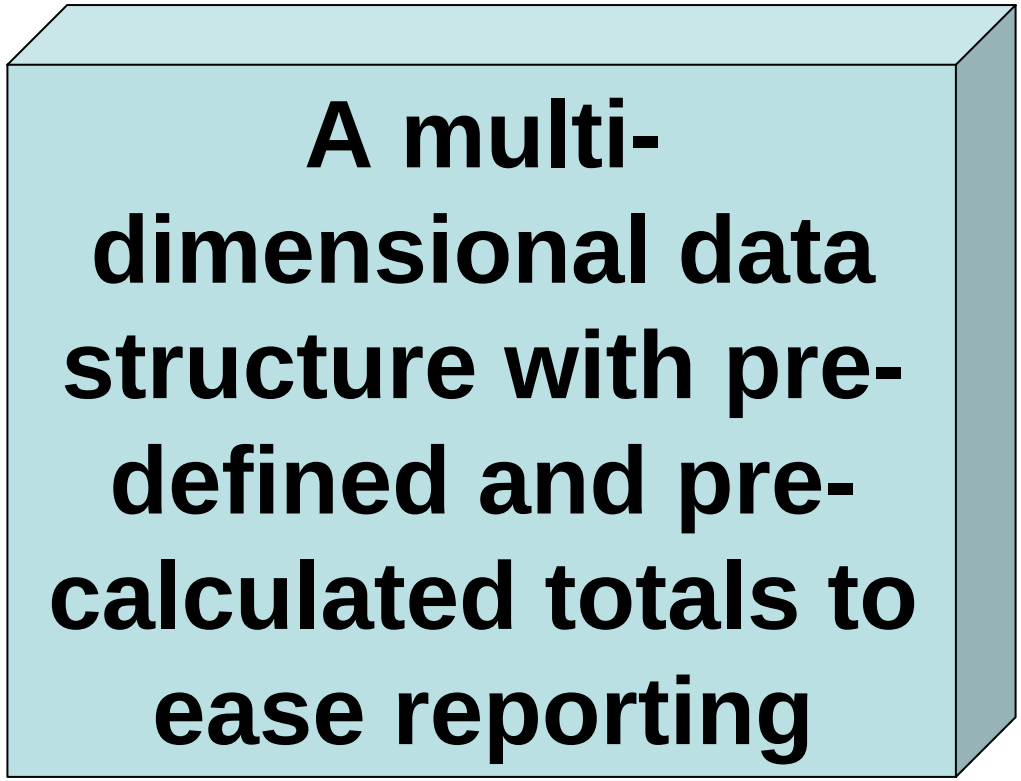
Demo Ouestion Three



34. It is now clear that Hispanic Females from Mexico make up the largest single population of High Risk Positives in 2004 in Tennessee.

What is a Cube?

What is a Cube?



What is a Cube?

Note – A data structure is a computer representation of data that is “seen” through software, in this case a web browser. For instance the actual structure of an Excel spreadsheet’s data and its representation are two separate things.

The Official Definition

"On-Line Analytical Processing (OLAP) is a way of organizing data to fit the way you analyze and manage it, so that less time and effort is needed to create your reports. When you create an OLAP cube from a query, you turn the flat set of records into a structured hierarchy, or cube, that allows reports to focus on the desired level of detail. You also predefine the summary values for the reports, which speeds up report calculation." – MS Product Literature

Pros and Cons

- Pros –
  - Inexpensive to implement
    - $0.00 if you have Microsoft Office and FrontPage
  - Friendly interface, easy to use for
    - Users
    - System Administrators and
    - Programmers

Pros and Cons

- Pros –
  - Much more efficient than spreadsheet or pivot table reporting
  - Easier to refresh with current data than other products
    - When users open a report, they always have the most current data in the Cube. Refreshing is not required.

Pros and Cons

- Cons –
  - Tempting to use even when not applicable
  - Can create an appearance of great accuracy that may not necessarily be true
  - Easy to reflect user biases in reporting
  - Very difficult to fully automate
    - Will always require some human interaction

Pros and Cons – A Warning

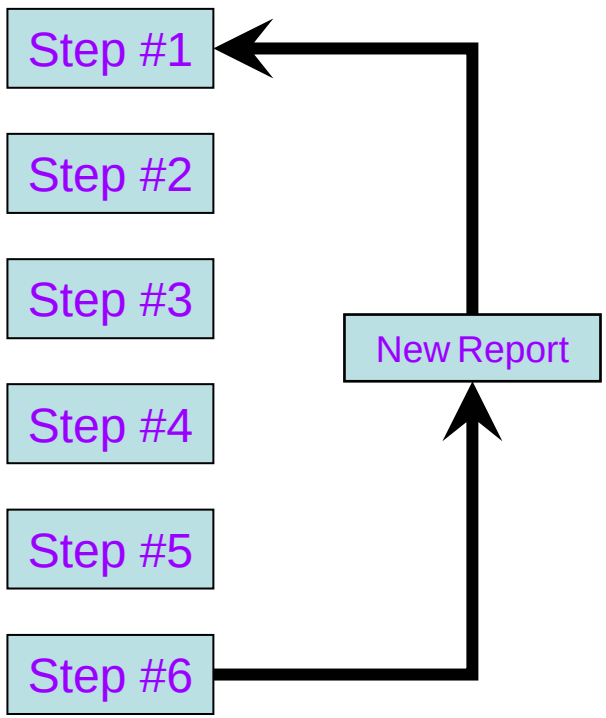
WARNING

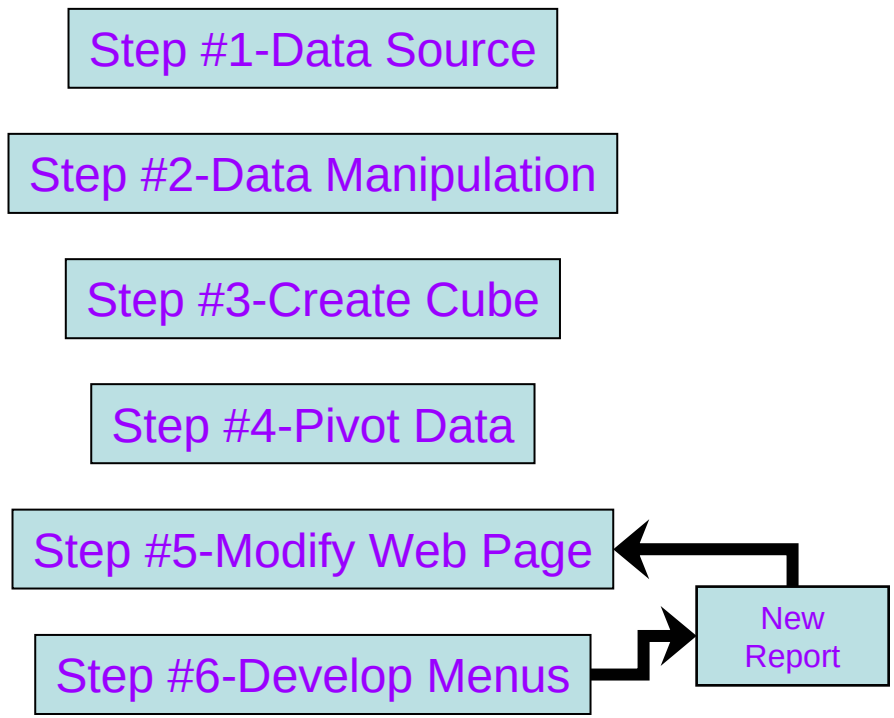
Since this product can put a lot of information on users desktop, think about your policies concerning who can provide data outside your agency and who or what should check the accuracy of that data prior to release!!!

Report Generation Paradigms

Old Paradigm – always start all over

New Paradigm – Just change the current visualization





Step 1: Data Source Identification and Research

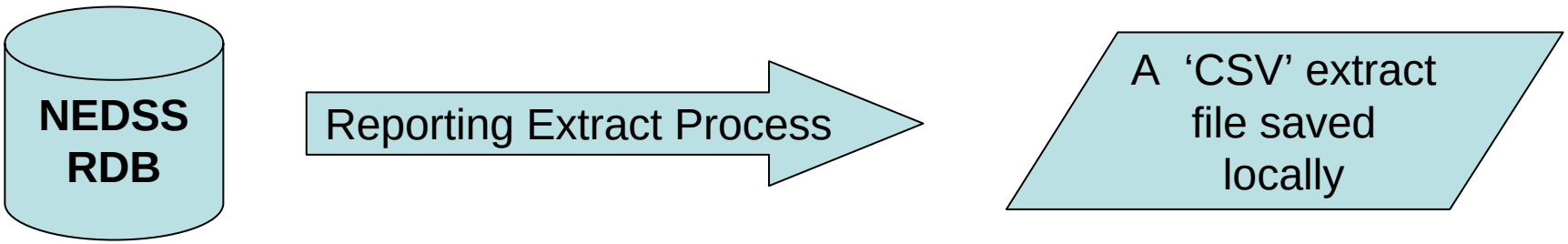
- Collect and document your metadata, i.e., gather facts about the data
- Compose meaningful data element names
- Define and understand data elements to be counted or summed
- Identify redundant data elements
- Identify related data elements (Jurisdiction Code and Jurisdiction Name)
- Eliminate unique individual identifiers

Step 1: Data Source Identification and Research

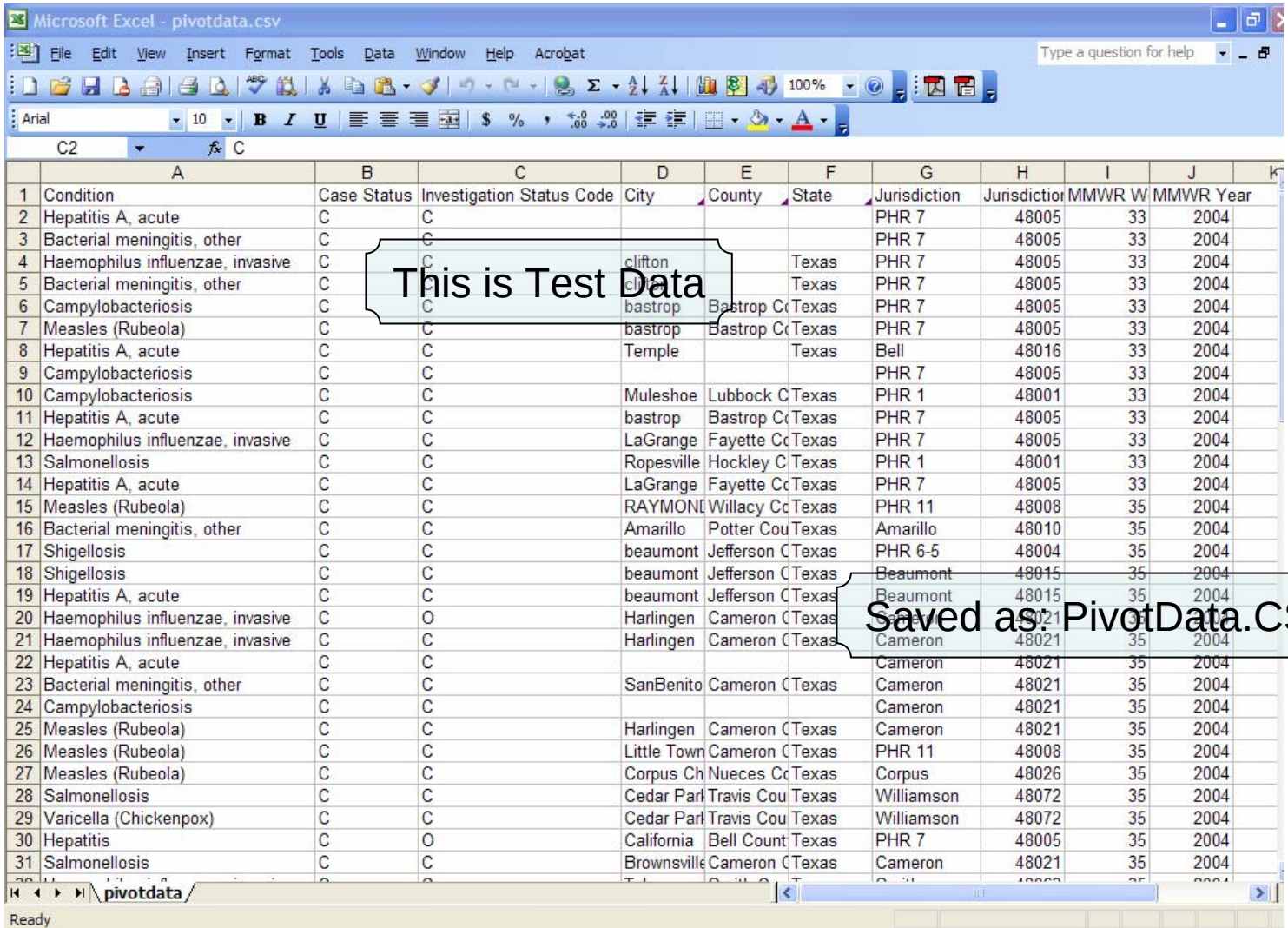
1. You will find that you will spend a great deal of time here studying and defining your data.

2. This can be almost anything: MS Access, MS Sql, MySql, Oracle, Dbase, MS Excel, a CSV file, or a comma delimited text file or anything for which you have a **ODBC** driver

Step 1: Data Store



Step 1: Data Store



	A	B	C	D	E	F	G	H	I	J
1	Condition	Case Status	Investigation Status Code	City	County	State	Jurisdiction	Jurisdictior	MMWR W	MMWR Year
2	Hepatitis A, acute	C	C				PHR 7	48005	33	2004
3	Bacterial meningitis, other	C	C				PHR 7	48005	33	2004
4	Haemophilus influenzae, invasive	C	C	clifton		Texas	PHR 7	48005	33	2004
5	Bacterial meningitis, other	C	C	clifton		Texas	PHR 7	48005	33	2004
6	Campylobacteriosis	C	C	bastrop	Bastrop C(	Texas	PHR 7	48005	33	2004
7	Measles (Rubeola)	C	C	bastrop	Bastrop C(	Texas	PHR 7	48005	33	2004
8	Hepatitis A, acute	C	C	Temple		Texas	Bell	48016	33	2004
9	Campylobacteriosis	C	C				PHR 7	48005	33	2004
10	Campylobacteriosis	C	C	Muleshoe	Lubbock C	Texas	PHR 1	48001	33	2004
11	Hepatitis A, acute	C	C	bastrop	Bastrop C(	Texas	PHR 7	48005	33	2004
12	Haemophilus influenzae, invasive	C	C	LaGrange	Fayette C(	Texas	PHR 7	48005	33	2004
13	Salmonellosis	C	C	Ropesville	Hockley C	Texas	PHR 1	48001	33	2004
14	Hepatitis A, acute	C	C	LaGrange	Fayette C(	Texas	PHR 7	48005	33	2004
15	Measles (Rubeola)	C	C	RAYMONI	Willacy C(	Texas	PHR 11	48008	35	2004
16	Bacterial meningitis, other	C	C	Amarillo	Potter Cou	Texas	Amarillo	48010	35	2004
17	Shigellosis	C	C	beaumont	Jefferson C	Texas	PHR 6-5	48004	35	2004
18	Shigellosis	C	C	beaumont	Jefferson C	Texas	Beaumont	48015	35	2004
19	Hepatitis A, acute	C	C	beaumont	Jefferson C	Texas	Beaumont	48015	35	2004
20	Haemophilus influenzae, invasive	C	O	Harlingen	Cameron C	Texas	Cameron	48021	35	2004
21	Haemophilus influenzae, invasive	C	C	Harlingen	Cameron C	Texas	Cameron	48021	35	2004
22	Hepatitis A, acute	C	C				Cameron	48021	35	2004
23	Bacterial meningitis, other	C	C	SanBenito	Cameron C	Texas	Cameron	48021	35	2004
24	Campylobacteriosis	C	C				Cameron	48021	35	2004
25	Measles (Rubeola)	C	C	Harlingen	Cameron C	Texas	Cameron	48021	35	2004
26	Measles (Rubeola)	C	C	Little Town	Cameron C	Texas	PHR 11	48008	35	2004
27	Measles (Rubeola)	C	C	Corpus Ch	Nueces Cc	Texas	Corpus	48026	35	2004
28	Salmonellosis	C	C	Cedar Parl	Travis Cou	Texas	Williamson	48072	35	2004
29	Varicella (Chickenpox)	C	C	Cedar Parl	Travis Cou	Texas	Williamson	48072	35	2004
30	Hepatitis	C	O	California	Bell Count	Texas	PHR 7	48005	35	2004
31	Salmonellosis	C	C	Brownsville	Cameron C	Texas	Cameron	48021	35	2004

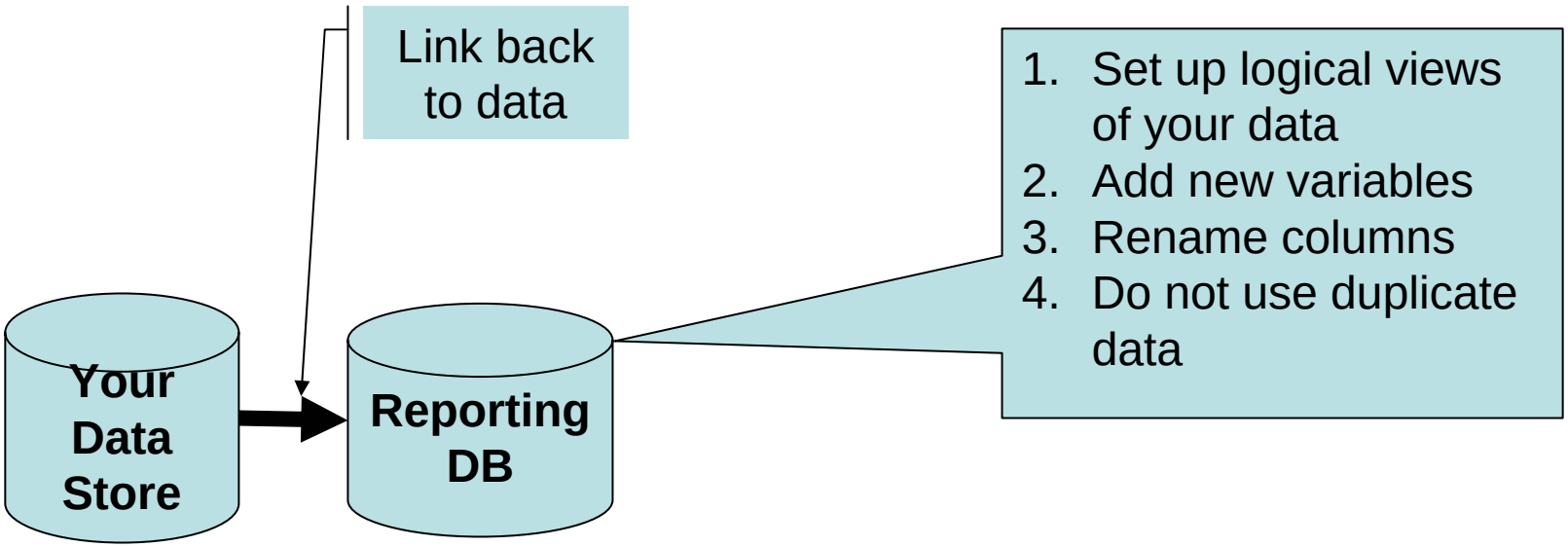
Step 2: Data Manipulation

- Supply meaningful data column names
- Eliminate redundant or identifier columns
- Add new variables – examples
  - If necessary, add a column purely for counting or summing
  - Join jurisdiction code and jurisdiction name
  - Subtract client DOB from the date of test to get client age at time of test
- This is were you lay the stage to make the data in the columns ‘pretty’

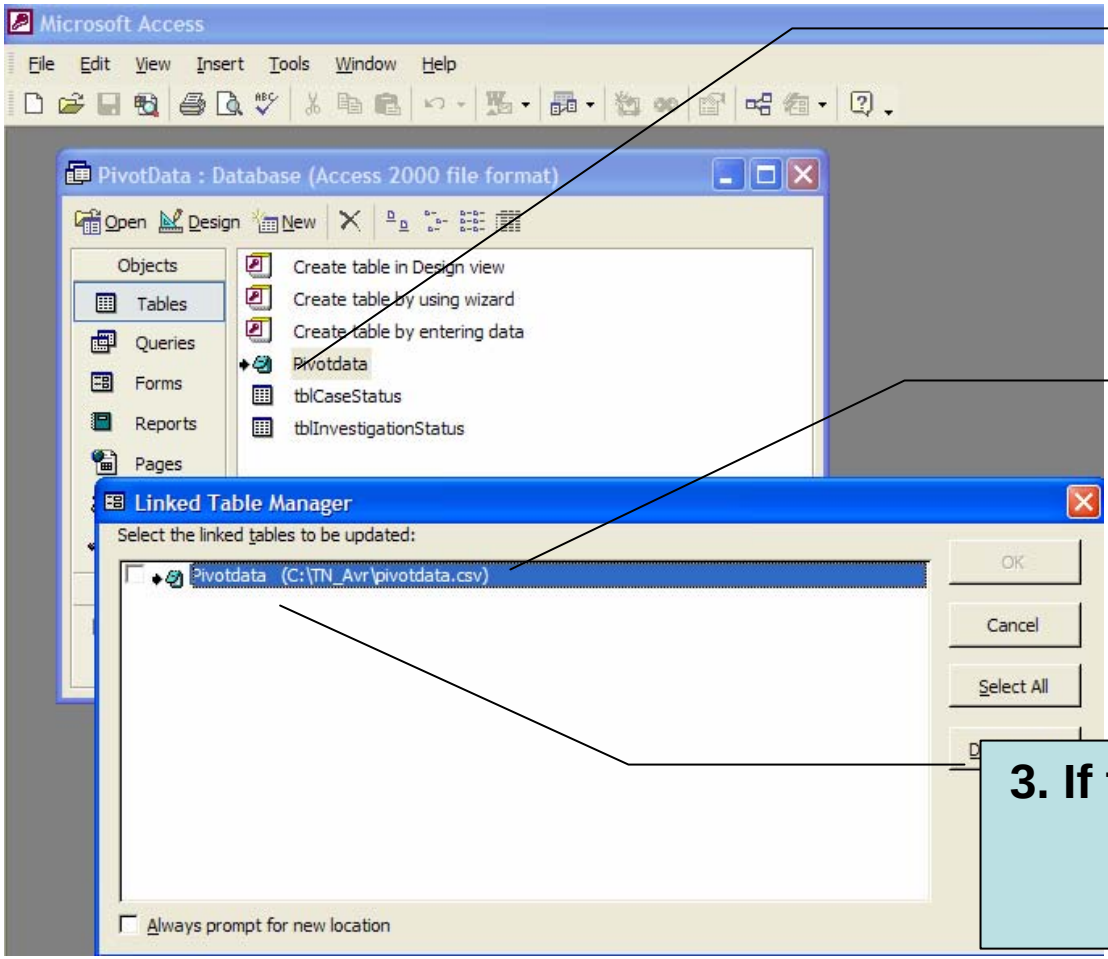
Step 2: Data Manipulation

Please note that the above process lets us transform the data from format found in the original source to a more **logical** and **useful** view of the data for later use.

Step 2: Data Manipulation



Step 2: Data Manipulation

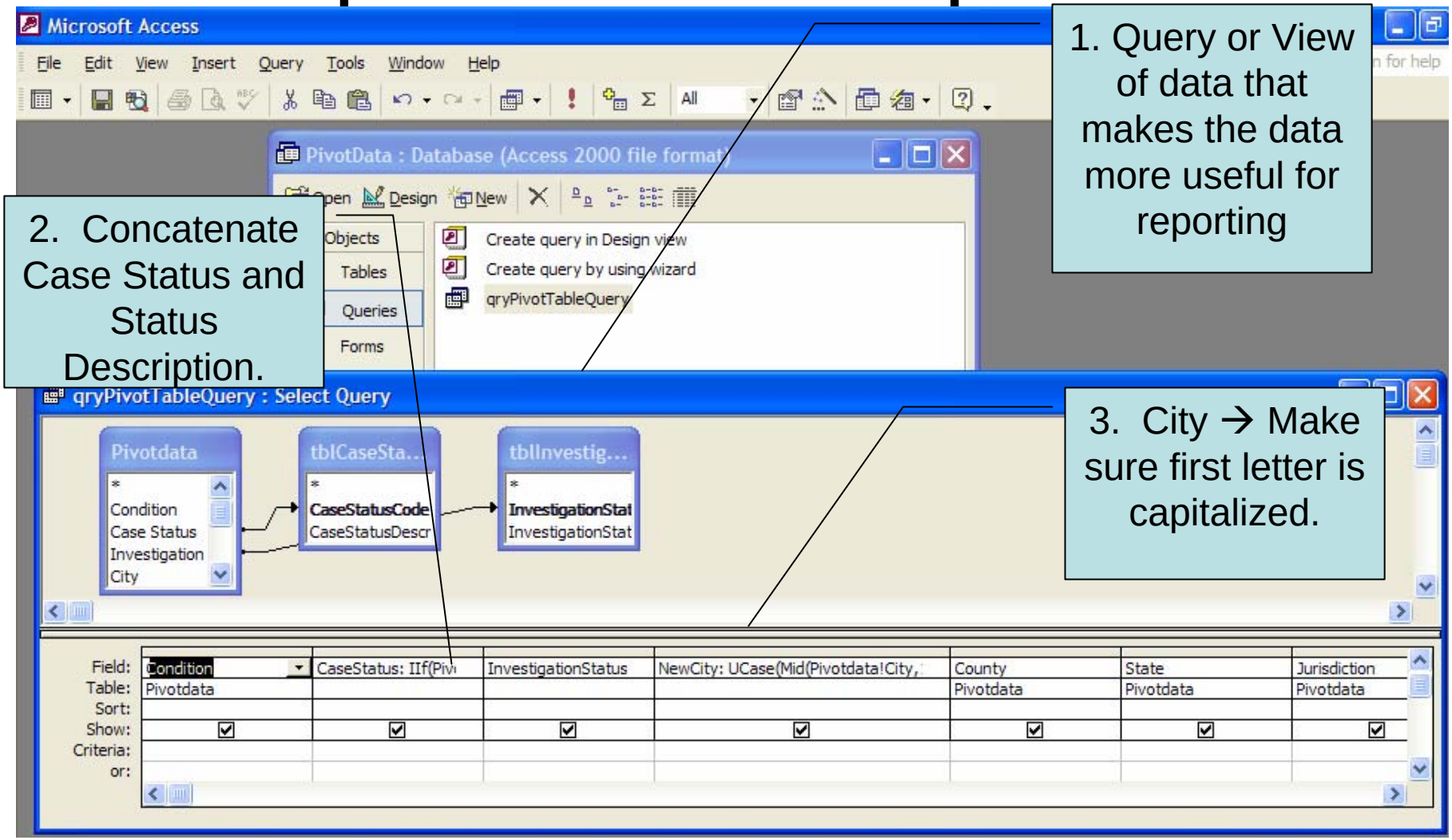


1. Our data table

2. Linked to the .CSV extract from our data source

3. If the underlying extract changes, the DB changes!!

Step 2: Data Manipulation

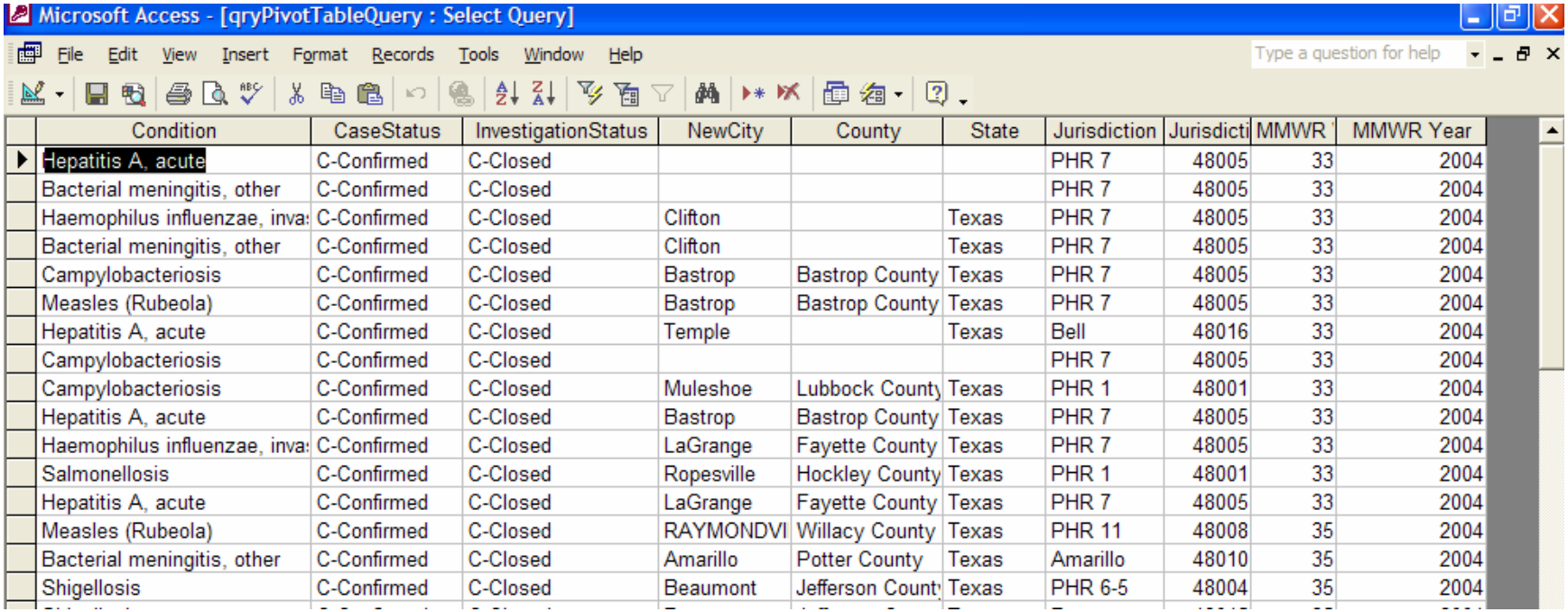


1. Query or View of data that makes the data more useful for reporting

2. Concatenate Case Status and Status Description.

3. City → Make sure first letter is capitalized.

Step 2: Data Manipulation



Condition	CaseStatus	InvestigationStatus	NewCity	County	State	Jurisdiction	Jurisdicti	MMWR	MMWR Year
Hepatitis A, acute	C-Confirmed	C-Closed				PHR 7	48005	33	2004
Bacterial meningitis, other	C-Confirmed	C-Closed				PHR 7	48005	33	2004
Haemophilus influenzae, inva	C-Confirmed	C-Closed	Clifton		Texas	PHR 7	48005	33	2004
Bacterial meningitis, other	C-Confirmed	C-Closed	Clifton		Texas	PHR 7	48005	33	2004
Campylobacteriosis	C-Confirmed	C-Closed	Bastrop	Bastrop County	Texas	PHR 7	48005	33	2004
Measles (Rubeola)	C-Confirmed	C-Closed	Bastrop	Bastrop County	Texas	PHR 7	48005	33	2004
Hepatitis A, acute	C-Confirmed	C-Closed	Temple		Texas	Bell	48016	33	2004
Campylobacteriosis	C-Confirmed	C-Closed				PHR 7	48005	33	2004
Campylobacteriosis	C-Confirmed	C-Closed	Muleshoe	Lubbock County	Texas	PHR 1	48001	33	2004
Hepatitis A, acute	C-Confirmed	C-Closed	Bastrop	Bastrop County	Texas	PHR 7	48005	33	2004
Haemophilus influenzae, inva	C-Confirmed	C-Closed	LaGrange	Fayette County	Texas	PHR 7	48005	33	2004
Salmonellosis	C-Confirmed	C-Closed	Ropesville	Hockley County	Texas	PHR 1	48001	33	2004
Hepatitis A, acute	C-Confirmed	C-Closed	LaGrange	Fayette County	Texas	PHR 7	48005	33	2004
Measles (Rubeola)	C-Confirmed	C-Closed	RAYMONDVI	Willacy County	Texas	PHR 11	48008	35	2004
Bacterial meningitis, other	C-Confirmed	C-Closed	Amarillo	Potter County	Texas	Amarillo	48010	35	2004
Shigellosis	C-Confirmed	C-Closed	Beaumont	Jefferson Count	Texas	PHR 6-5	48004	35	2004

NOTE:

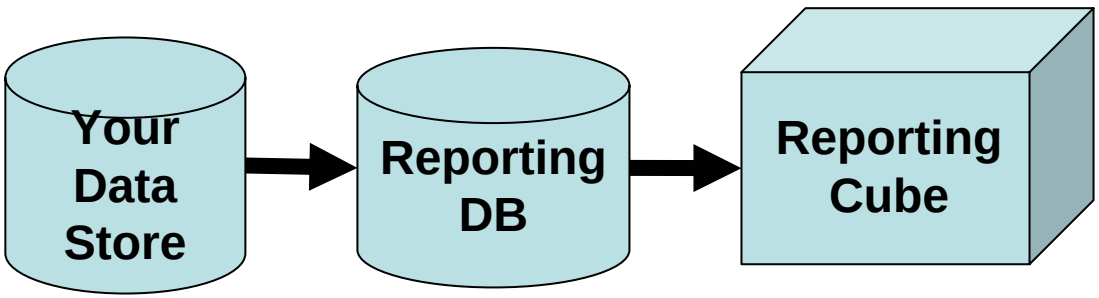
THIS IS THE CONSTRUCT THAT WILL BE USED TO BUILD OUR OLAP CUBE

Step 3: Create the Cube

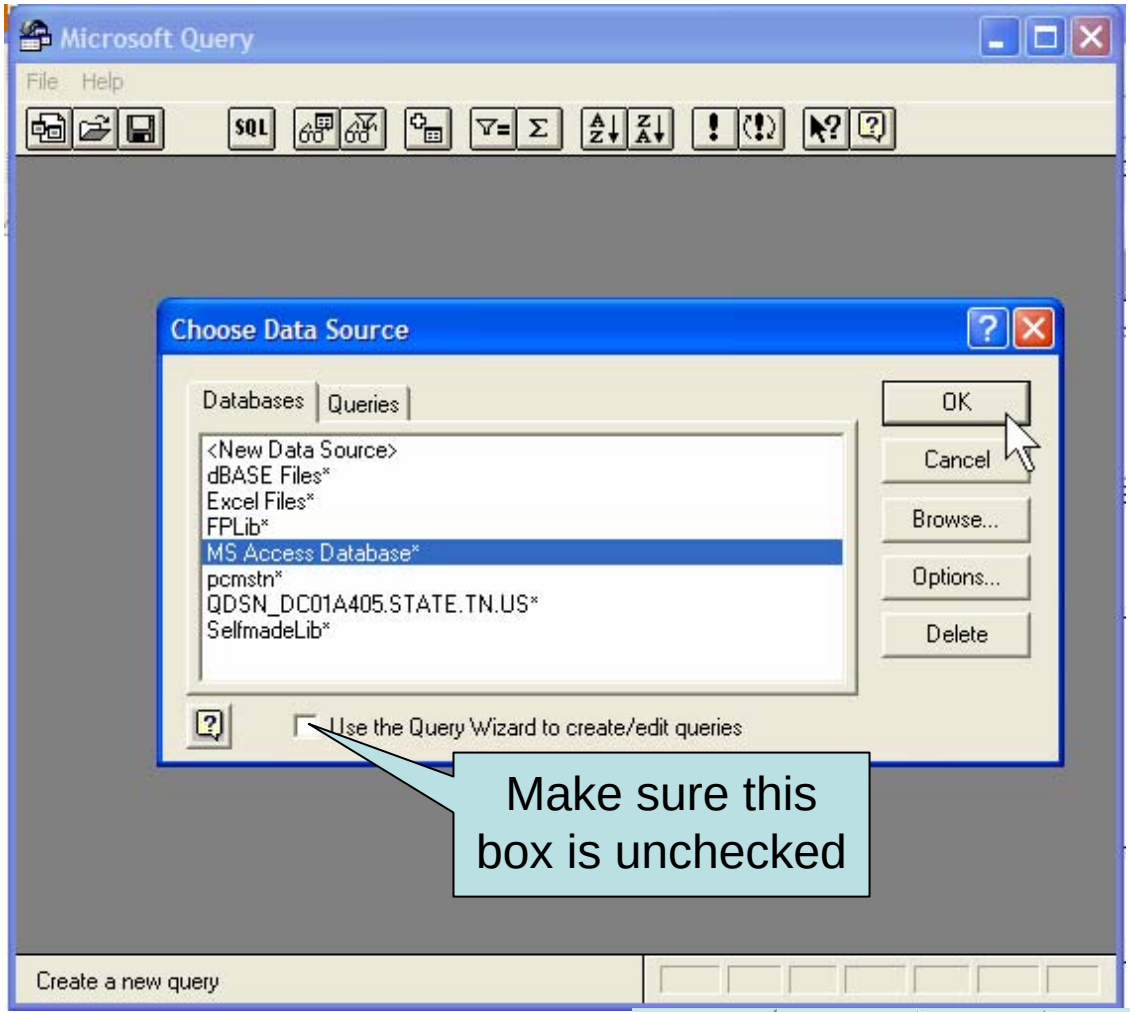
- Start MSQRY32.exe (find here: <drive>:\Program Files\Microsoft Office)
- Access the logical view you just created
- Make sure you have identified the focus columns for the reporting structure
- Save as an OLAP Cube

MSQRY32.EXE

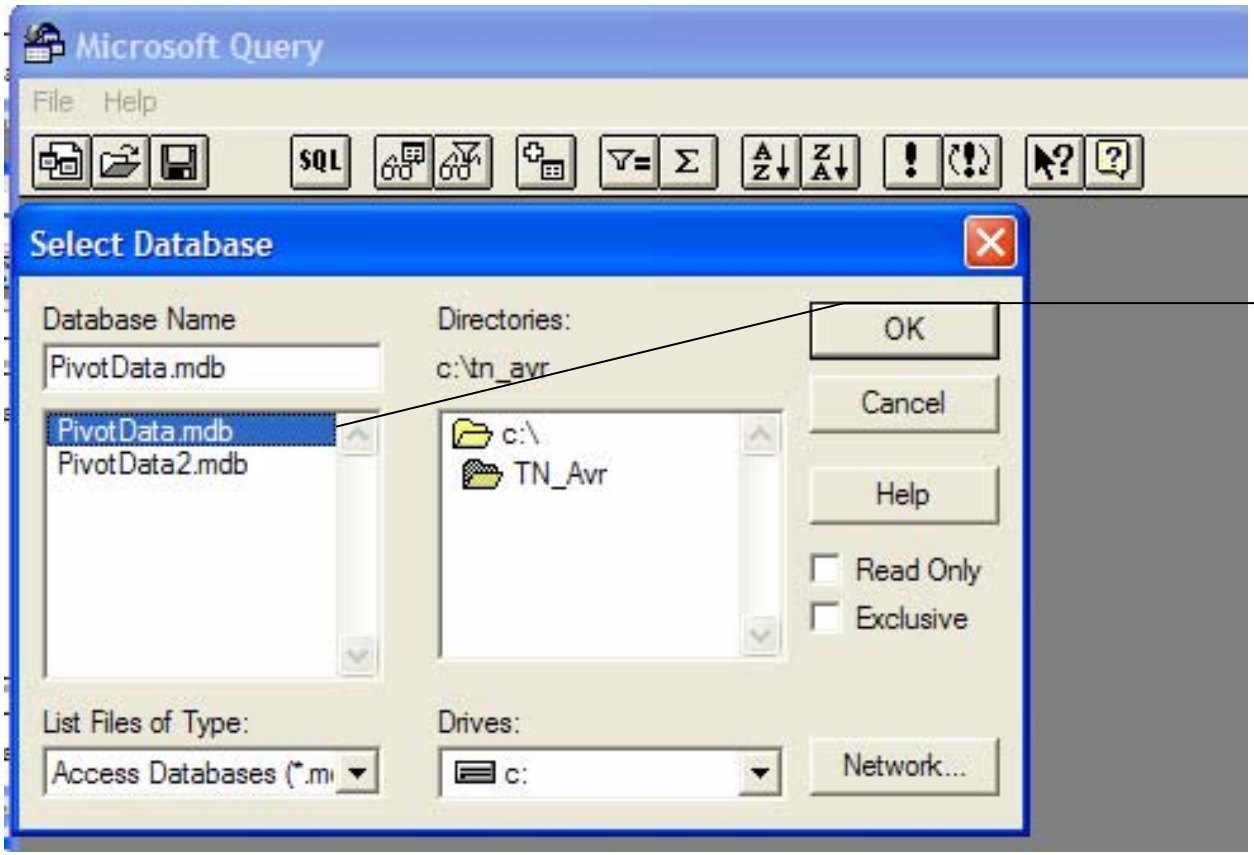
Step 3: Create the Cube



Step 3: Create the Cube

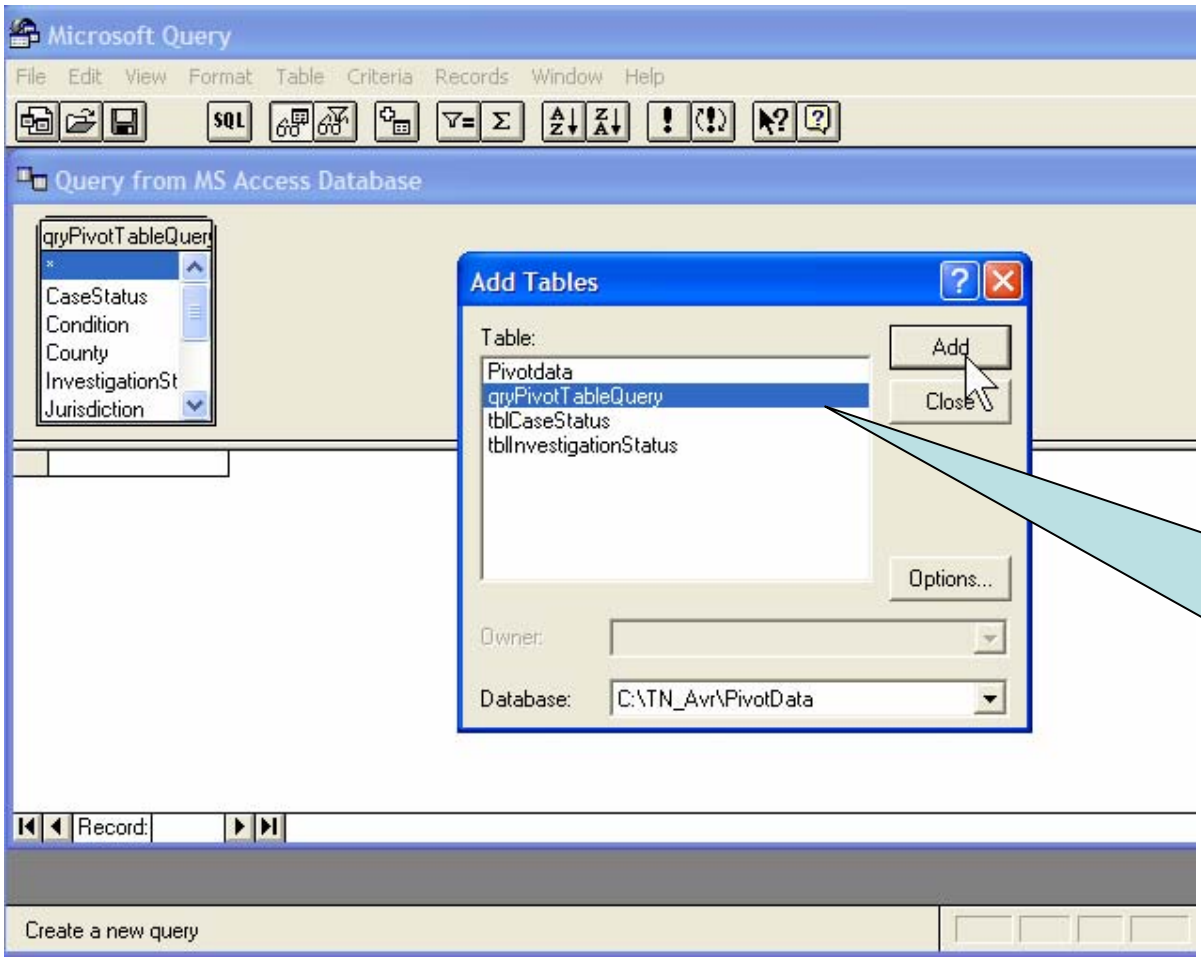


Step 3: Create the Cube



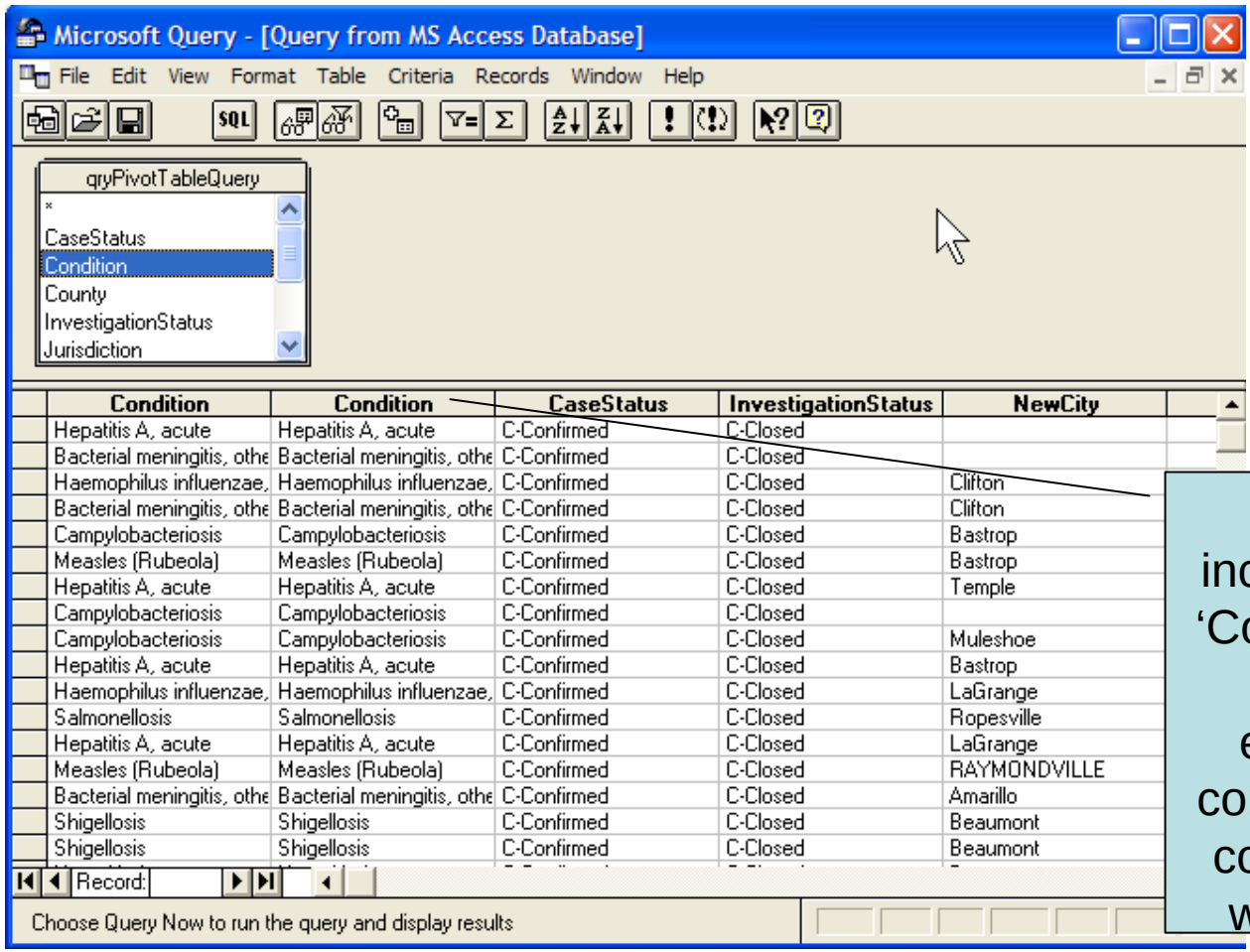
The database we just worked on

Step 3: Create the Cube



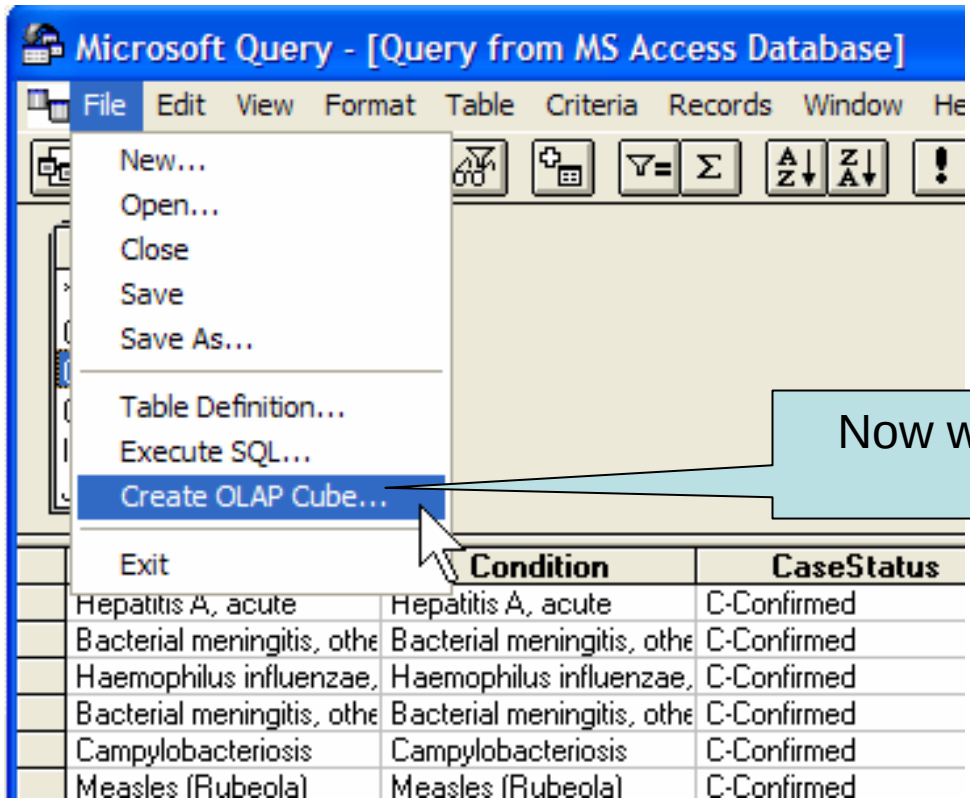
Pick the view of the data you created in the previous step

Step 3: Create the Cube



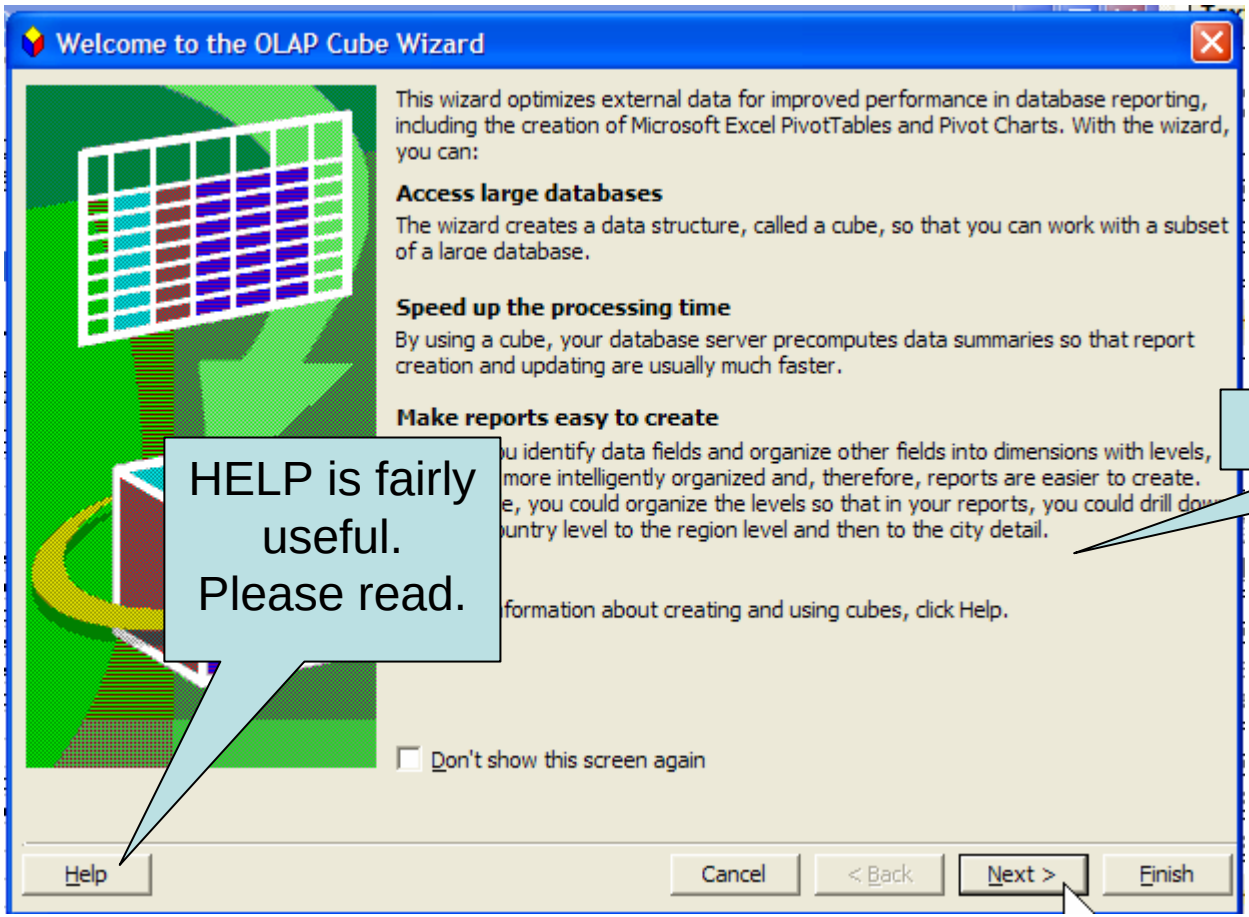
NOTE: Two incidents of the field ‘Condition’. One will be used to enumerate each condition and one to count the incidents within a condition

Step 3: Create the Cube

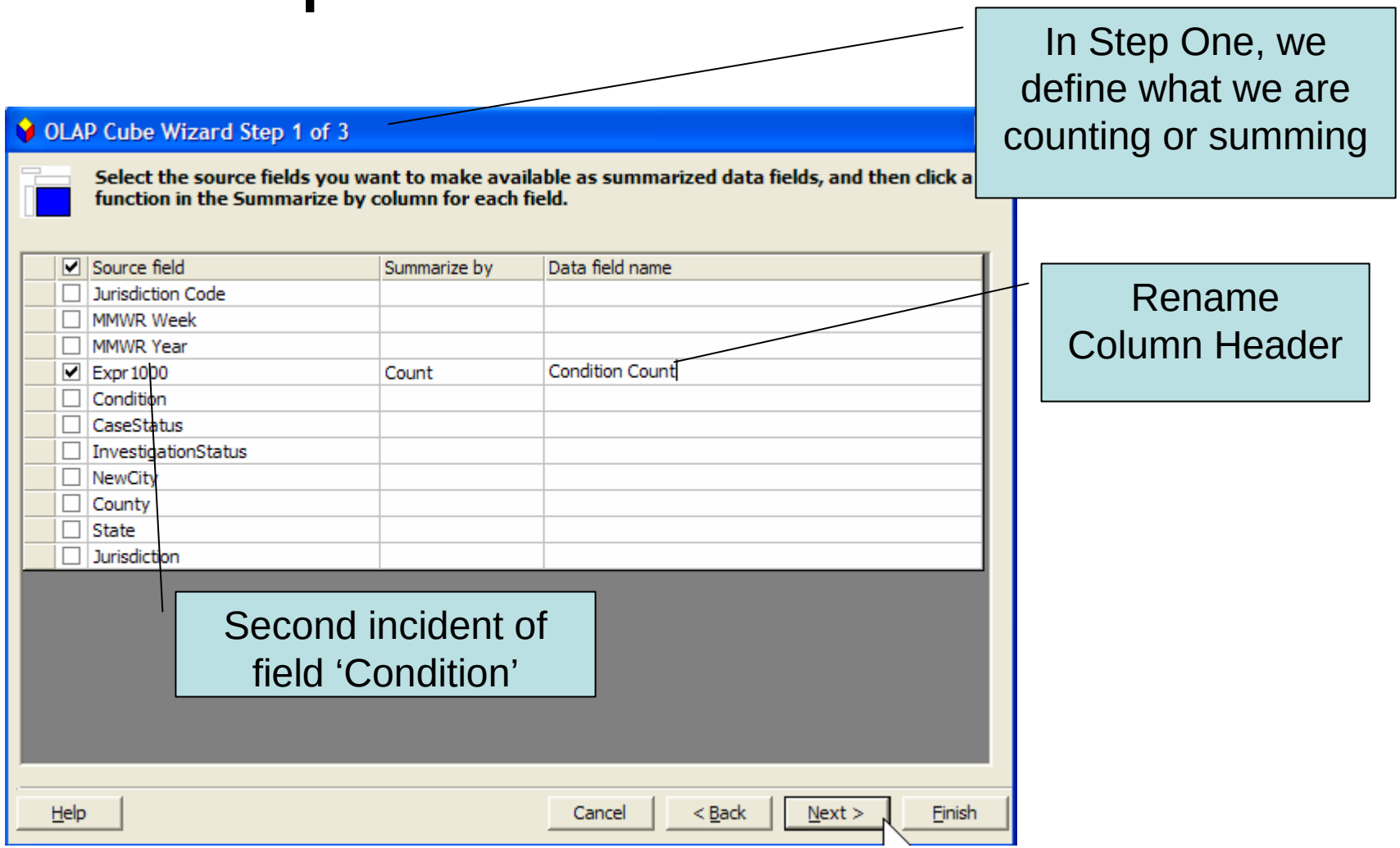


Now we create our OLAP Cube

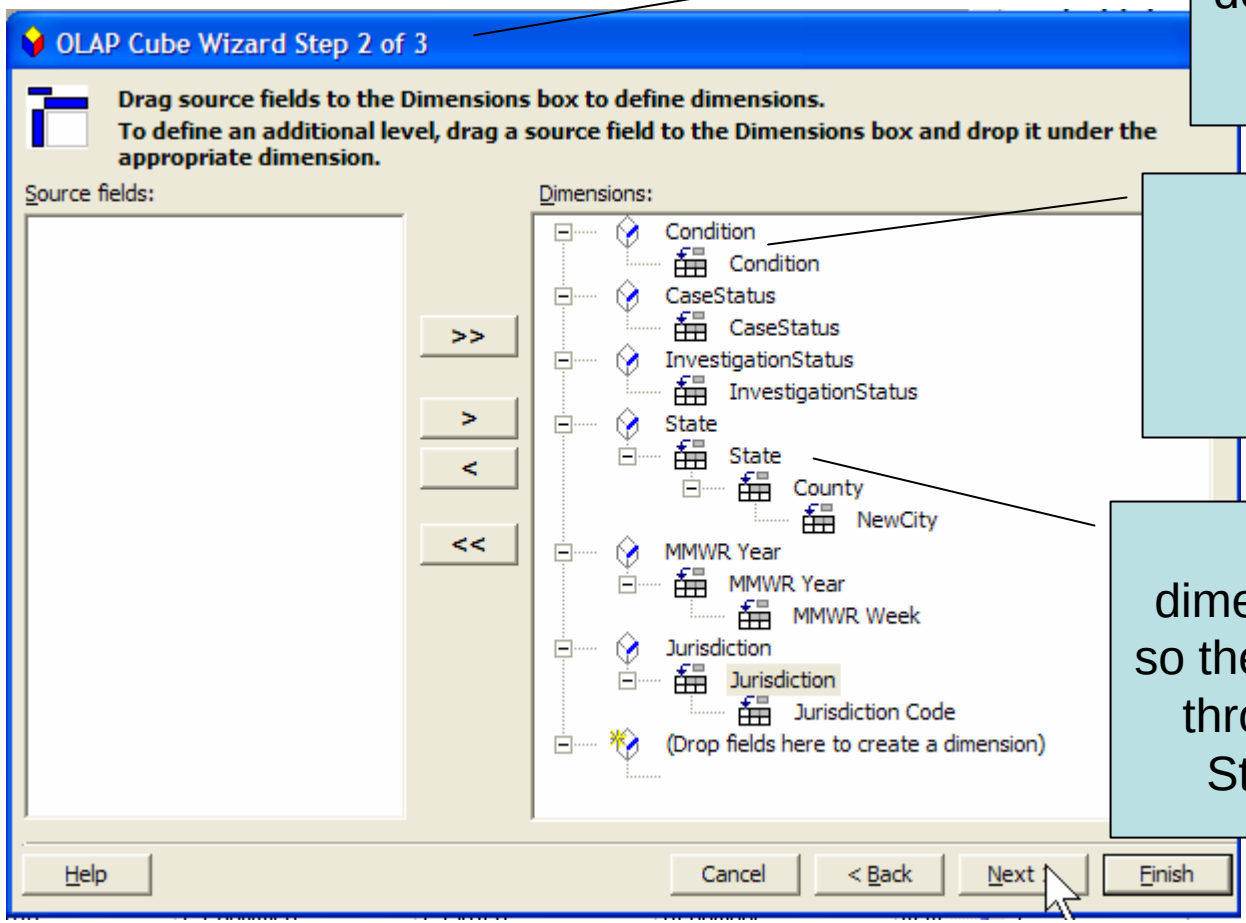
Step 3: Create the Cube



Step 3: Create the Cube



Step 3: Create the Cube



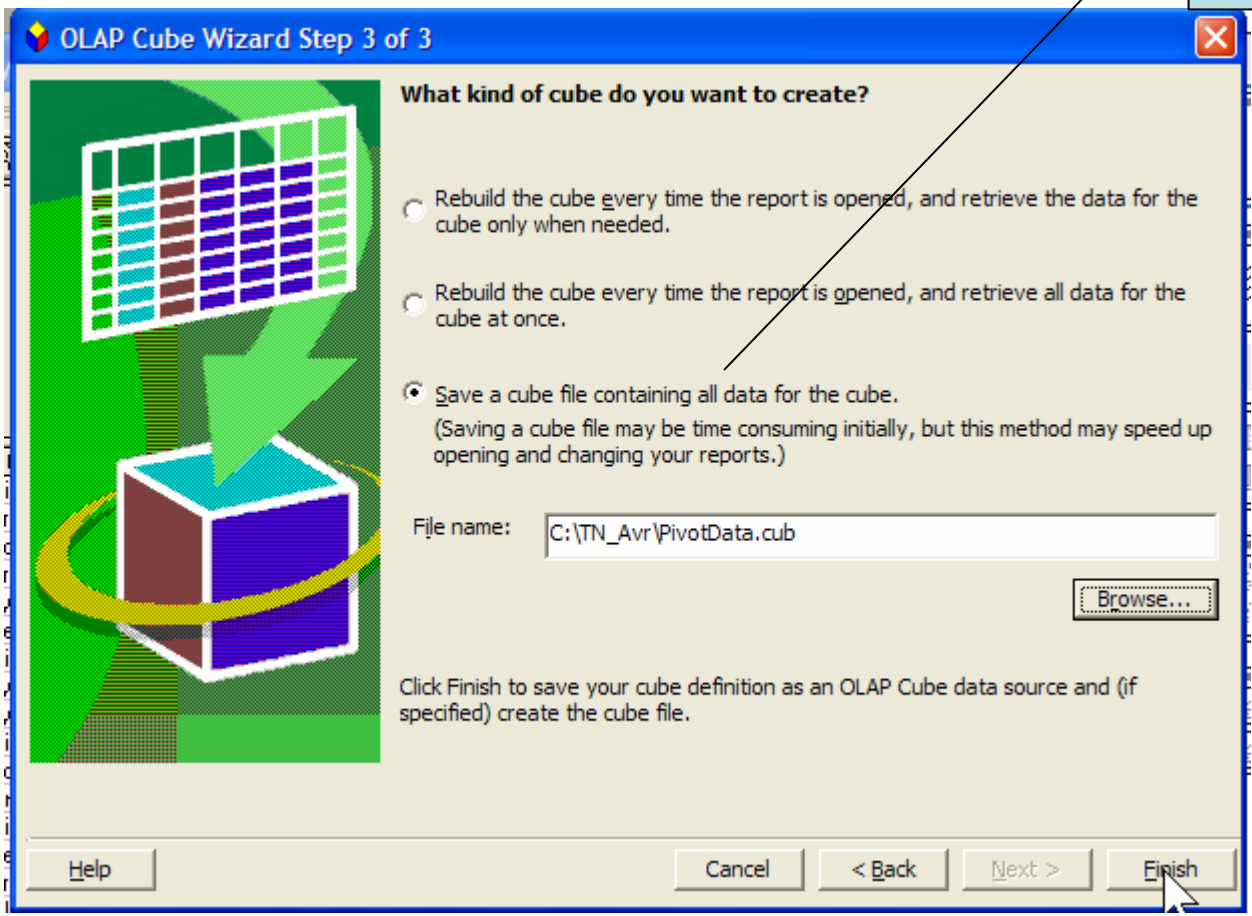
In Step Two, we define the dimensions of the report

Note that some dimensions are standalone, like 'Condition'

Note that some dimensions are stacked so the user can drill down through the data, like State, County, City

Step 3: Create the Cube

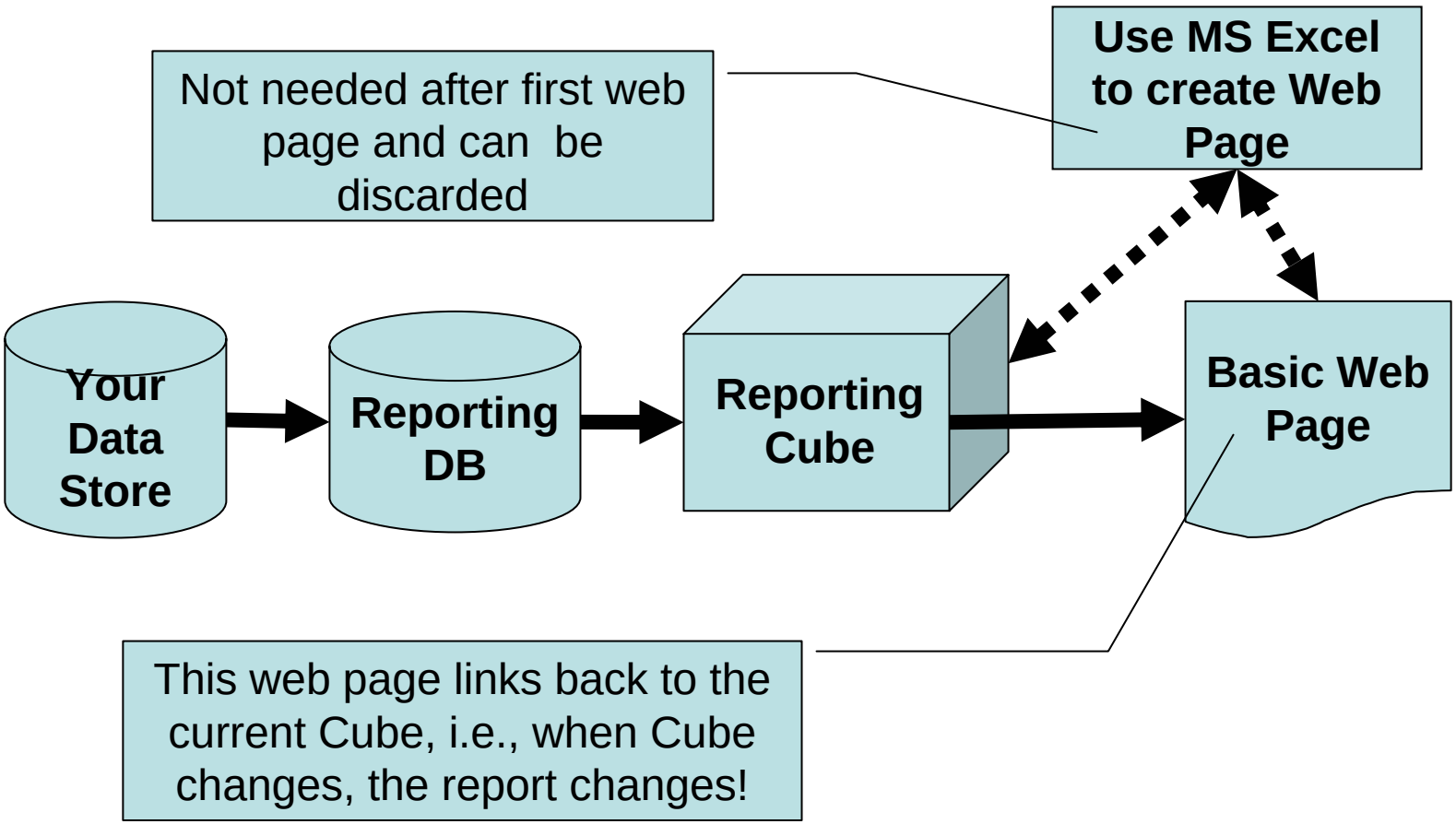
Offline Cube



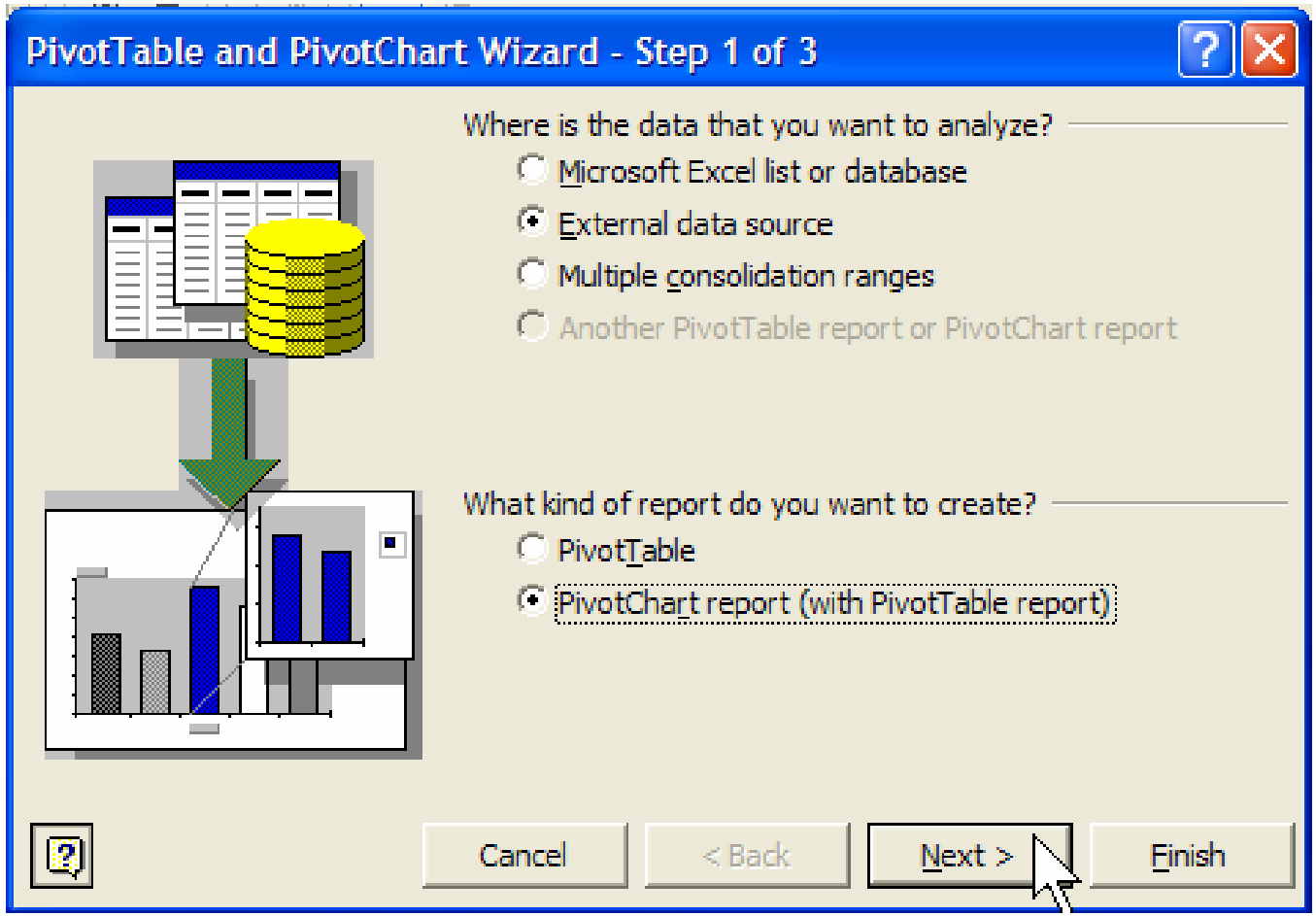
Step 4: Pivot the Data

- Use MS Excel
  - Use external data source
  - Create a PivotChart (with PivotTable report)
  - Save the PivotChart (NOT the Pivot Table!) as a web page
  - Close Excel and discard the Excel pivot table

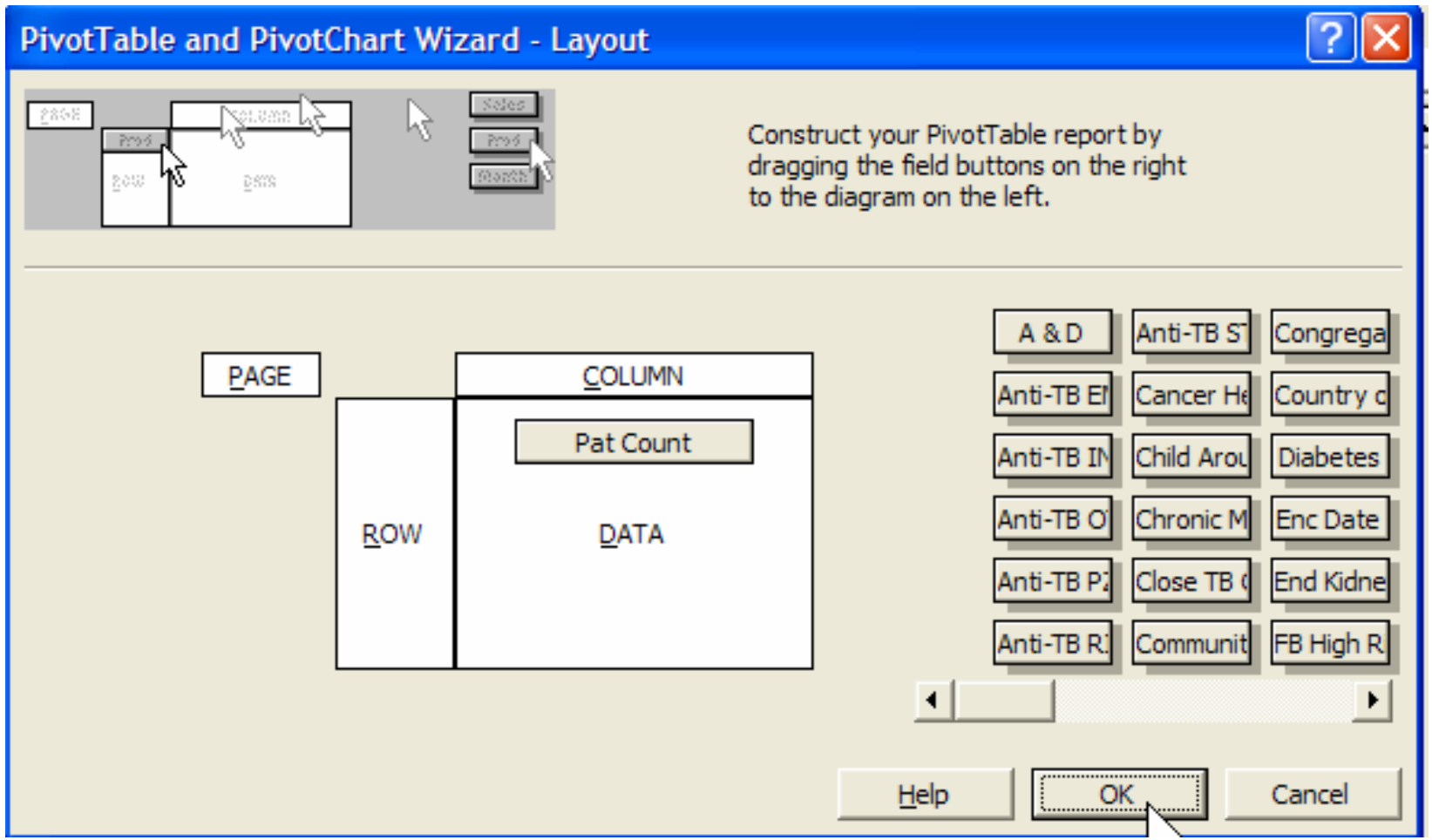
Step 4: Pivot the Data



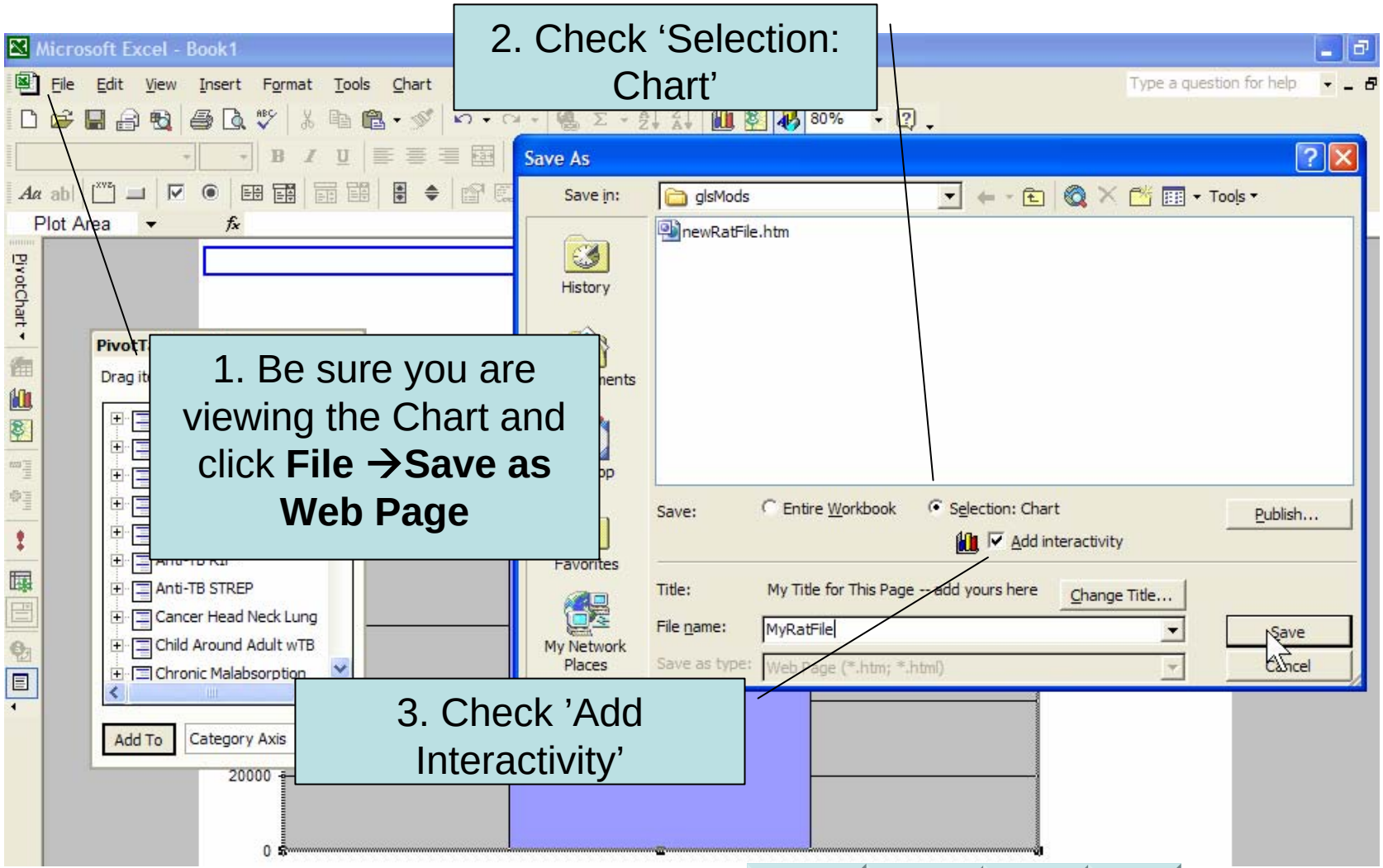
Step 4: Pivot the Data



Step 4: Pivot the Data



Step 4: Pivot the Data



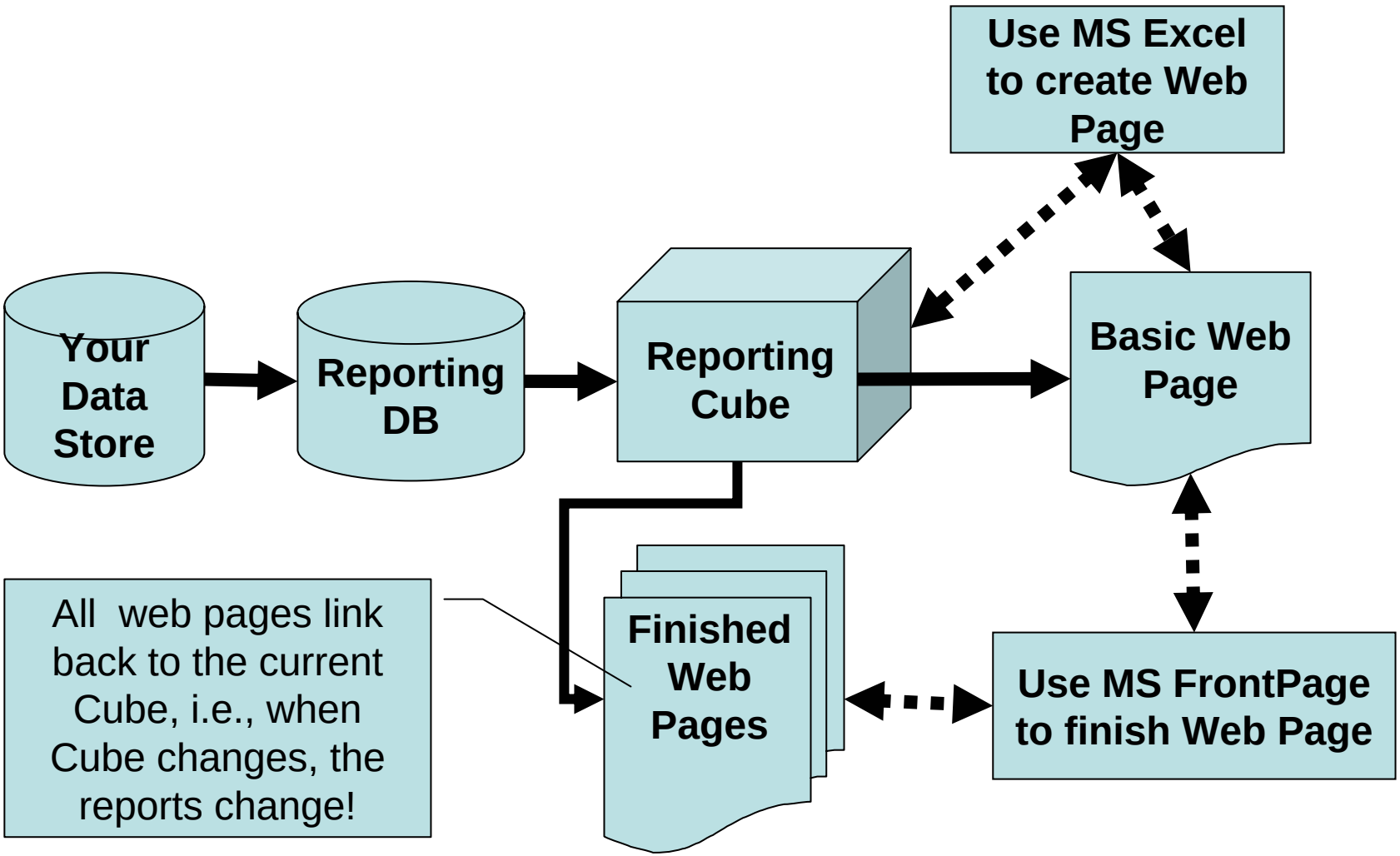
Step 5: Modify Web Page

- Use MS FrontPage to open the web page you just created
  - For the geeks: this exposes all the objects for these report structures
  - Build one report for instant gratification and to visualize the changes to follow
  - Add metadata (see next slide)
  - Finish starter set of reports

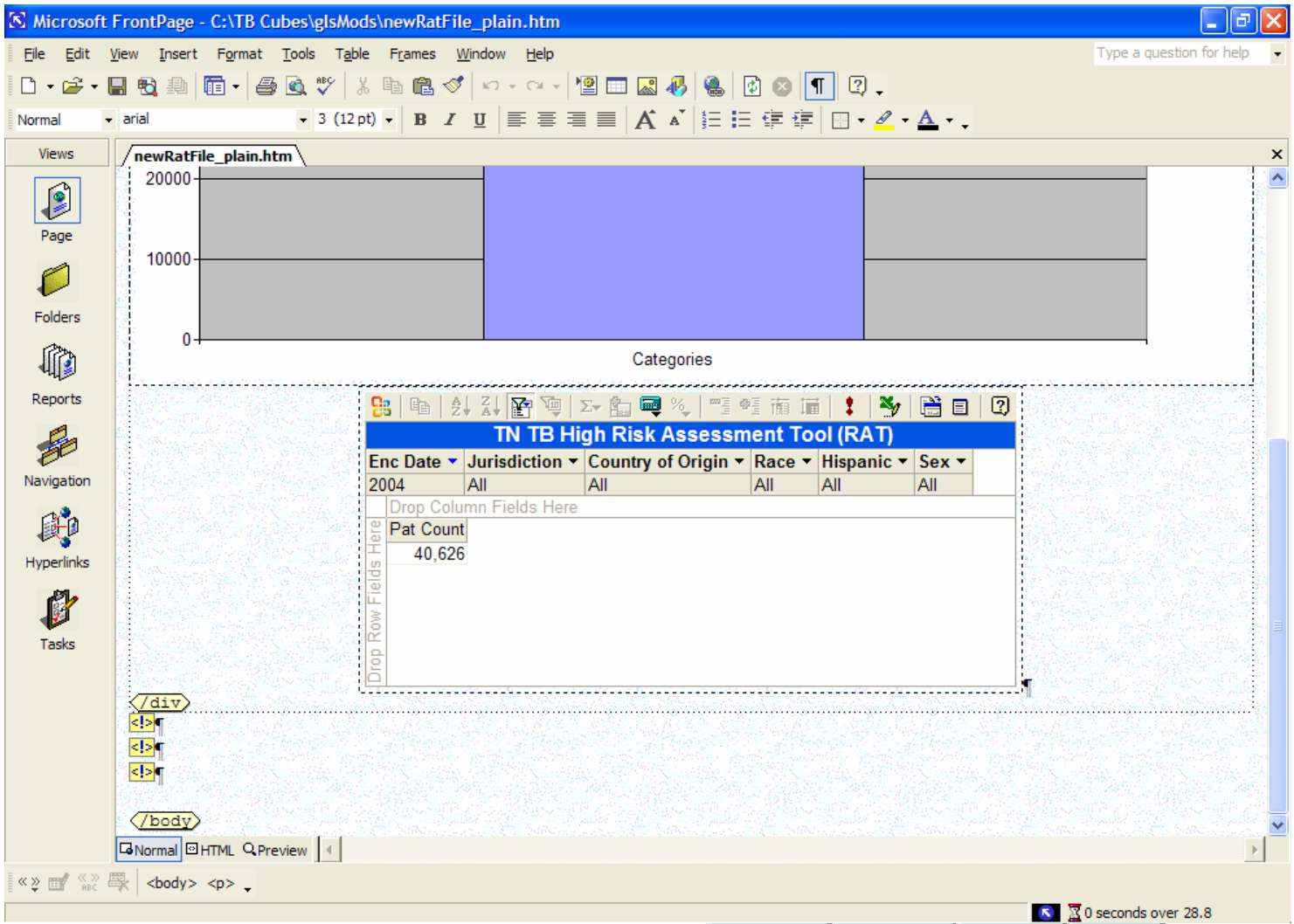
Step 5: Modify Web Page

- Make changes to make the data more **logical** and **useful**
  - Rename columns
  - Group data values, e.g. <blank>, “Not Entered” and “Unknown” into a new value
  - Add new calculated fields
  - Test as you go!

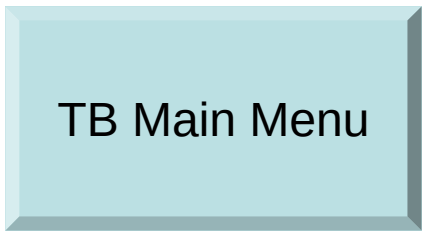
Step 5: Modify Web Page



Step 5: Modify Web Page

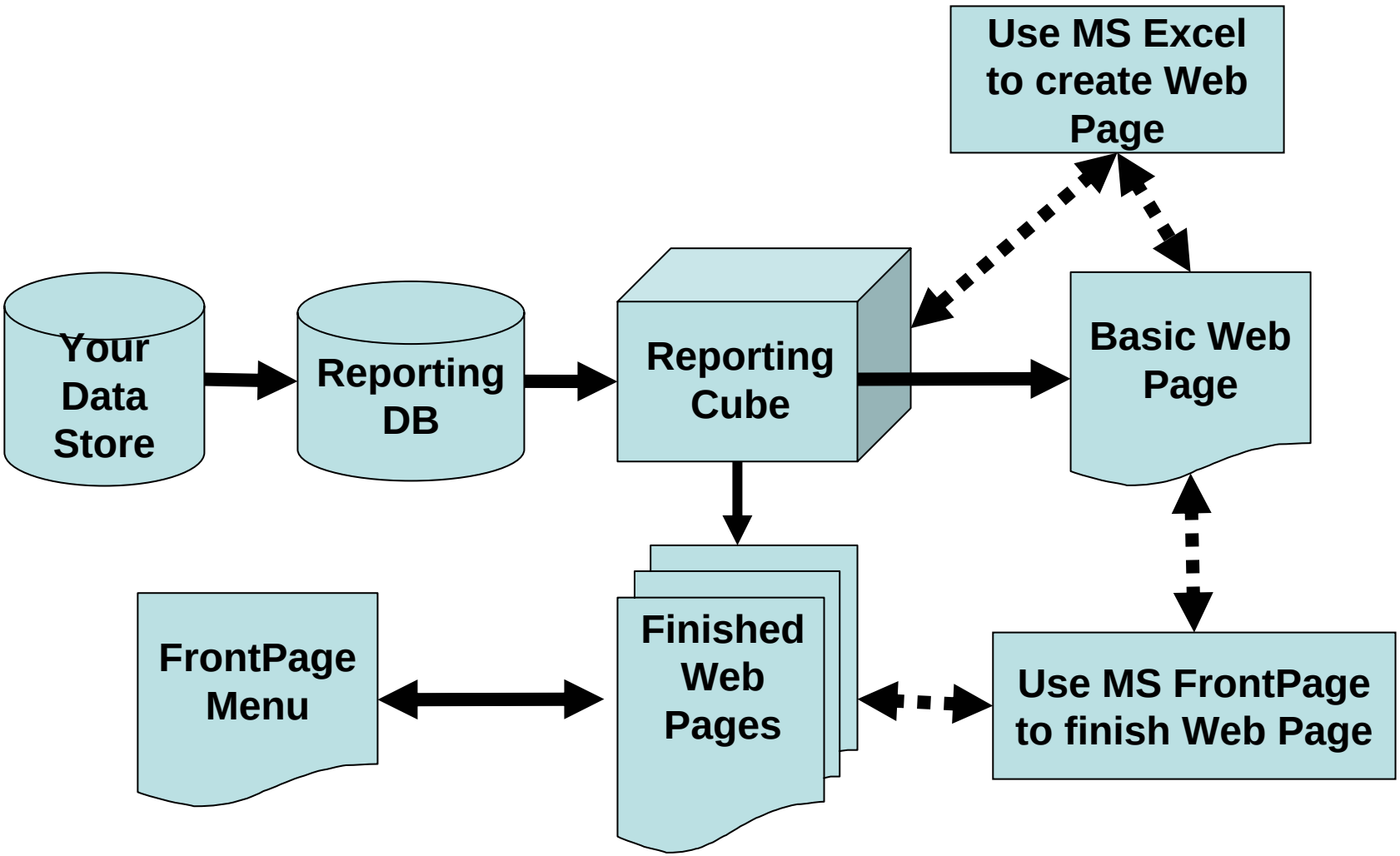


Step 6: Develop a Menu Page



Note: Two versions of many reports; one for percents (see % sign) and the other for counts (see # sign)

Step 6: Develop a Menu Page



Step 7: Marketing & Distribution

- Plan
- Train users
  - End users – push buttons and get reports
  - Power users – push buttons to get started building the report they need
  - System Administrators
    - Supply updated data in future
    - Train new users
    - Answer questions

Step 7: Marketing & Distribution

- Turn over to system administrators who provide current data and distribute to users
- It is implemented!

Be ready for quick round of modifications as power users define useful reports that need to be added to menus

Remember

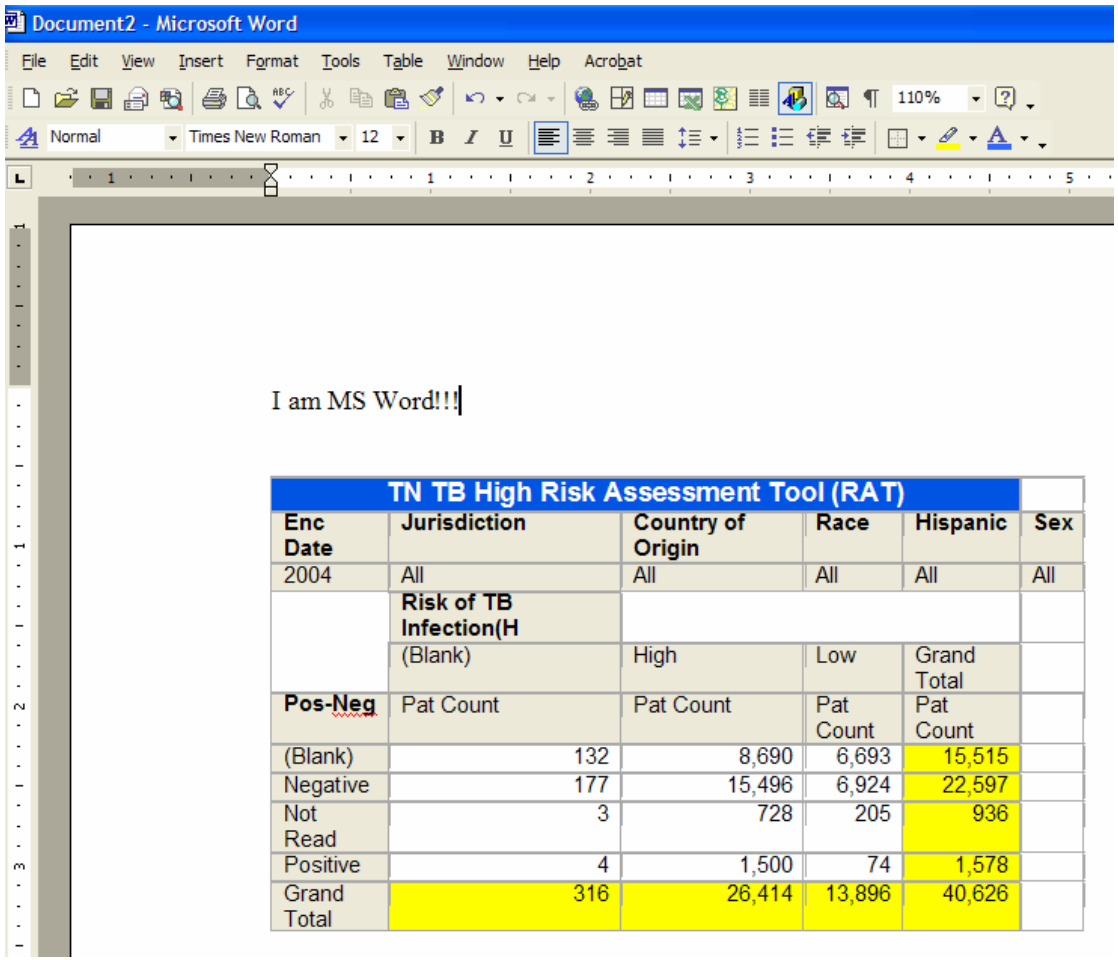
- To add new reports, you do not have to go back to the beginning.
- Go back to the Step 5 (MS FrontPage)
  - Modify a report
  - Save the new report
  - Add new report to a menu
  - Redistribute through your system administrator at next data refresh
  - Rock and roll

Gee Whiz Features

– Cut and Paste into Excel or Word

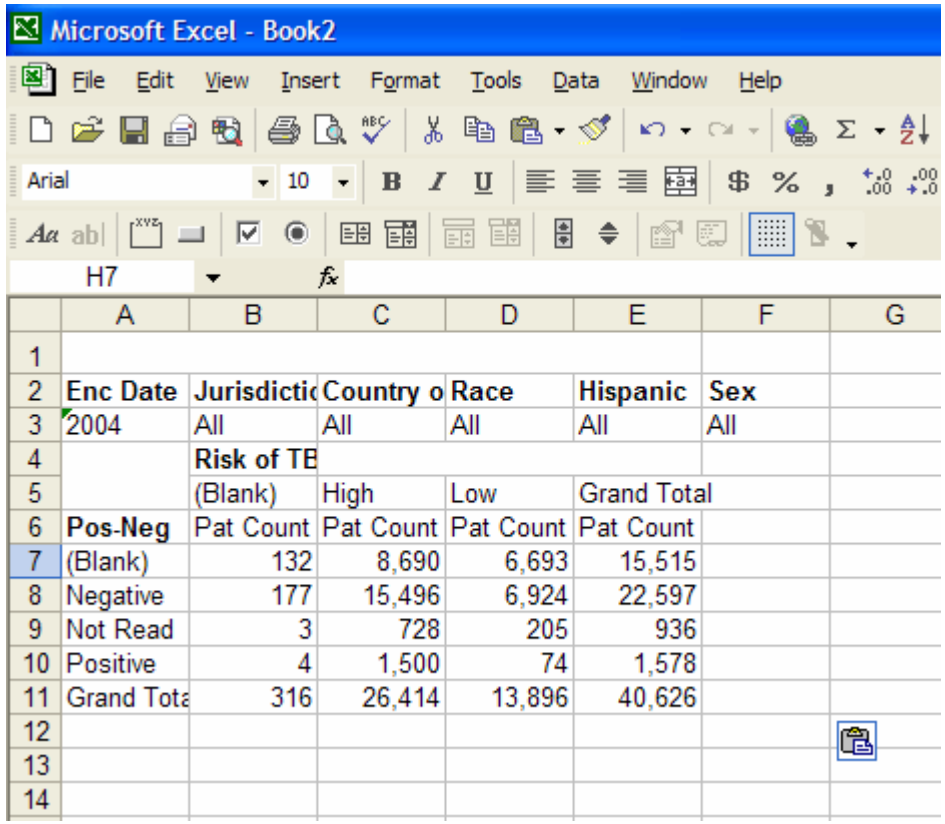
Report

Cut and Paste into Word



TN TB High Risk Assessment Tool (RAT)					
Enc Date	**Jurisdiction**	**Country of Origin**	**Race**	**Hispanic**	**Sex**
2004	All	All	All	All	All
	Risk of TB Infection(H				
	(Blank)	High	Low	Grand Total	
Pos-Neg	Pat Count	Pat Count	Pat Count	Pat Count	
(Blank)	132	8,690	6,693	15,515	
Negative	177	15,496	6,924	22,597	
Not Read	3	728	205	936	
Positive	4	1,500	74	1,578	
Grand Total	316	26,414	13,896	40,626	

Cut and Paste into Excel



	A	B	C	D	E	F	G
1							
2	**Enc Date**	**Jurisdictic**	**Country o**	**Race**	**Hispanic**	**Sex**	
3	2004	All	All	All	All	All	
4		**Risk of TB**					
5		(Blank)	High	Low	Grand Total		
6	**Pos-Neg**	Pat Count	Pat Count	Pat Count	Pat Count		
7	(Blank)	132	8,690	6,693	15,515		
8	Negative	177	15,496	6,924	22,597		
9	Not Read	3	728	205	936		
10	Positive	4	1,500	74	1,578		
11	Grand Tota	316	26,414	13,896	40,626		
12							
13							
14							

Gee Whiz Features

- Classic 2 x 2 tables

Gee Whiz Features

- Classic 2 x 2 tables
- Toggle filter to show differences between a subset of data and all the data

Counts and a Percent of Whatever

Counts

TN TB High Risk Assess			
Enc Date			
2004			
	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive	83	1491	1574
Negative	1297	21123	22420
Grand Total	1380	22614	23994

Row Percents

TN TB High Risk Asse			
Enc Date			
2004			
	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive	5.27%	94.73%	100.00%
Negative	5.79%	94.21%	100.00%
Grand Total	5.75%	94.25%	100.00%

Column Percents

TN TB High Risk Asses			
Enc Date			
2004			
	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive	6.01%	6.59%	6.56%
Negative	93.99%	93.41%	93.44%
Grand Total	100.00%	100.00%	100.00%

Percent of Grand Total

TN TB High Risk Asse			
Enc Date			
2004			
	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive	0.35%	6.21%	6.56%
Negative	5.41%	88.03%	93.44%
Grand Total	5.75%	94.25%	100.00%

Toggling the Filter

Filtered Group

TN TB High Risk Asses

Enc Date	Race
2004	Asian

	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive		31.25%	31.25%
Negative	0.37%	68.38%	68.75%
Grand Total	0.37%	99.63%	100.00%

Total Population

TN TB High Risk Asses

Enc Date	Race
2004	All

	HIV Mult Sex Partners		
	Y	N	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count
Positive	0.35%	6.21%	6.56%
Negative	5.41%	88.03%	93.44%
Grand Total	5.75%	94.25%	100.00%

Two Filtered Groups and Total

TN TB High Risk Assessment Tool (RAT)

Enc Date
2004

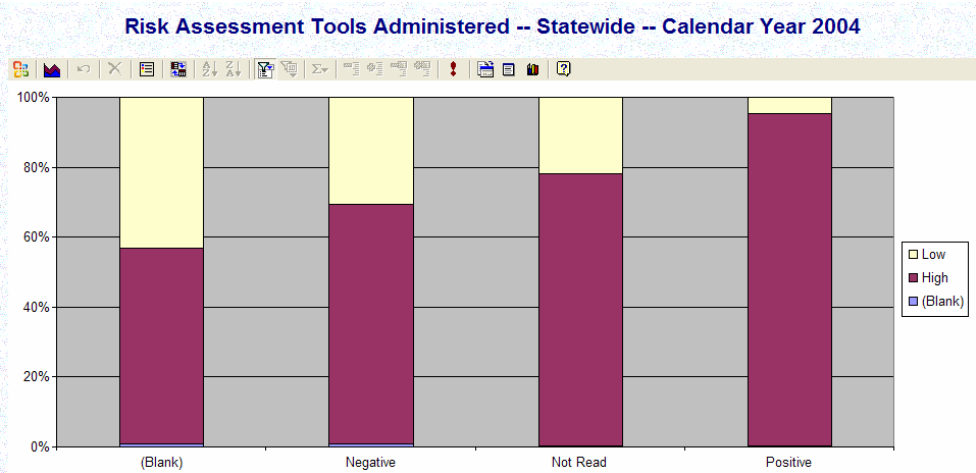
	Races	HIV Mult Sex Partners						
	⊟ Asian				⊟ Non Asian			Grand Total
	Y	N	Total		Y	N	Total	
Pos-Neg	Pat Count	Pat Count	Pat Count		Pat Count	Pat Count	Pat Count	Pat Count
Positive		31.37%	31.25%		6.02%	6.29%	6.28%	6.56%
Negative	100.00%	68.63%	68.75%		93.98%	93.71%	93.72%	93.44%
Grand Total	100.00%	100.00%	100.00%		100.00%	100.00%	100.00%	100.00%

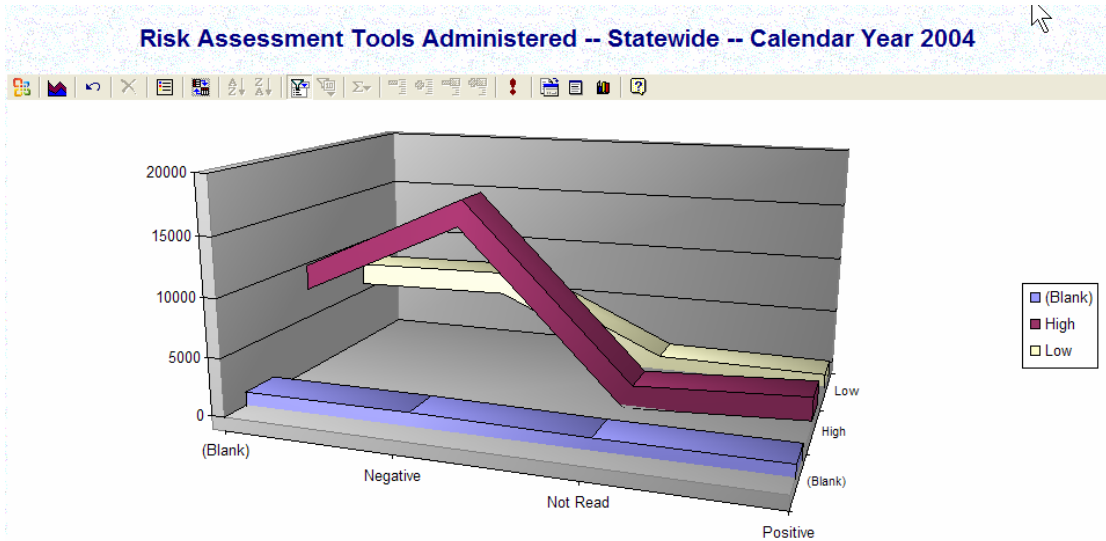
Gee Whiz Features

- Change chart format

TB

Change chart format



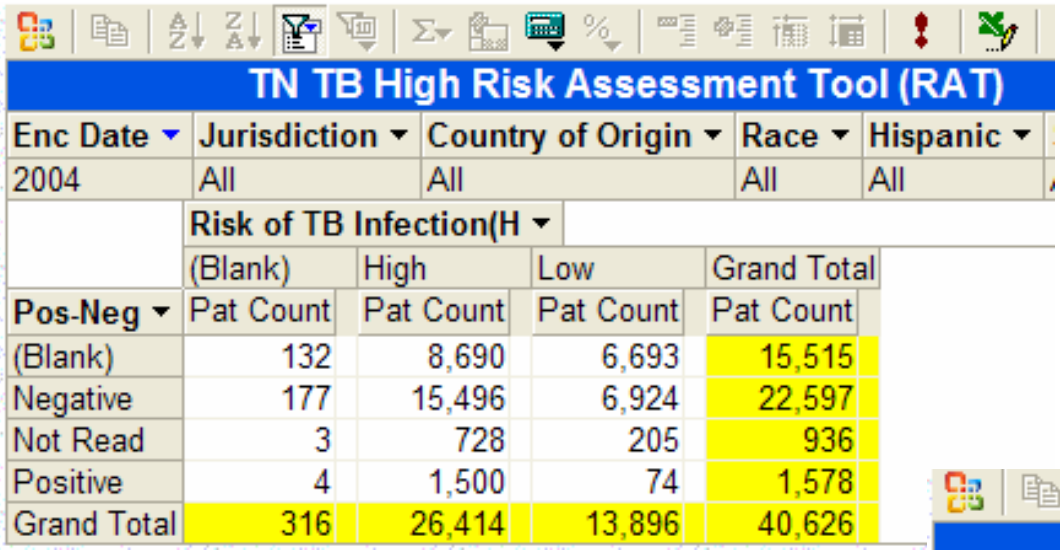


Gee Whiz Features

- Improve readability by decoding reported data values, e.g., “H” becomes “High Risk”
- Alter sequence of values

TB

Readabilty & Sequence of Values



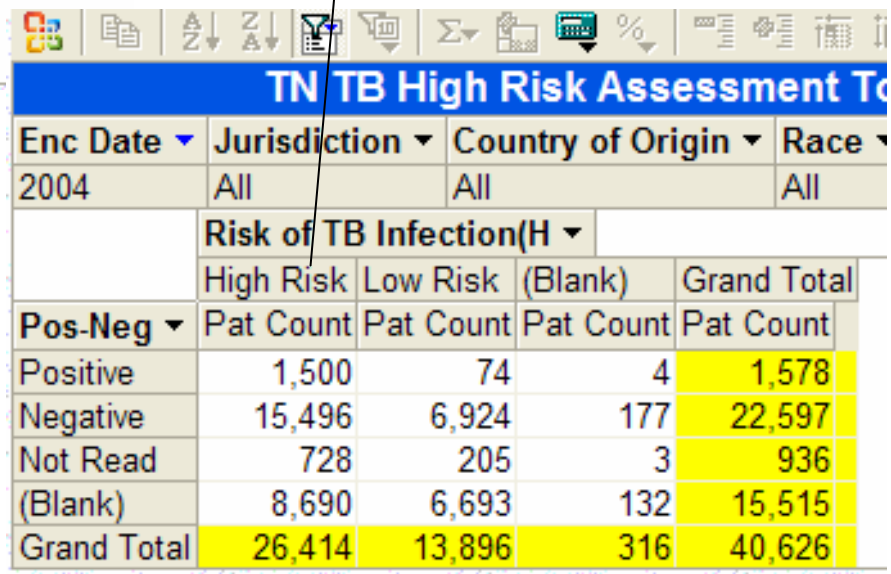
TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Country of Origin	Race	Hispanic
2004	All	All	All	All

	Risk of TB Infection(H			
	(Blank)	High	Low	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count	Pat Count
(Blank)	132	8,690	6,693	15,515
Negative	177	15,496	6,924	22,597
Not Read	3	728	205	936
Positive	4	1,500	74	1,578
Grand Total	316	26,414	13,896	40,626

With Value Renamed

With Values Rearranged



TN TB High Risk Assessment To

Enc Date	Jurisdiction	Country of Origin	Race
2004	All	All	All

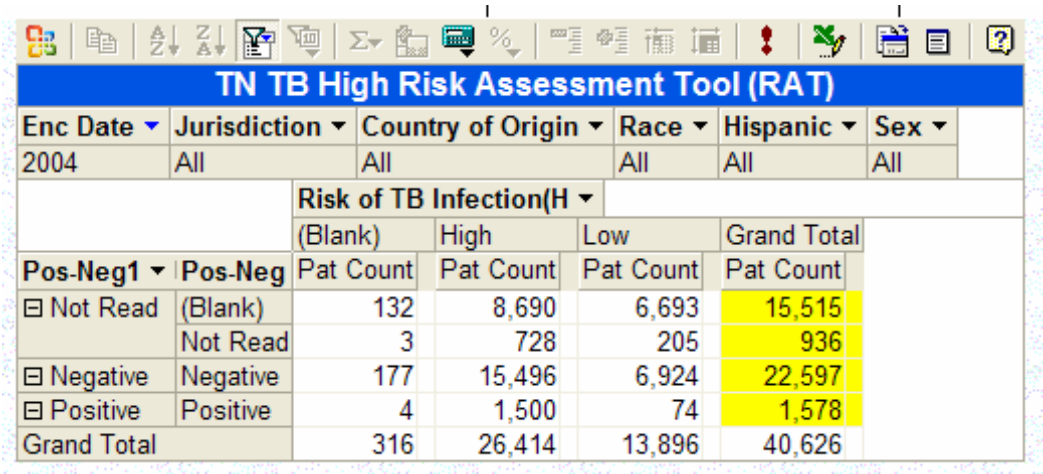
	Risk of TB Infection(H			
	High Risk	Low Risk	(Blank)	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count	Pat Count
Positive	1,500	74	4	1,578
Negative	15,496	6,924	177	22,597
Not Read	728	205	3	936
(Blank)	8,690	6,693	132	15,515
Grand Total	26,414	13,896	316	40,626

Gee Whiz Features

TB

- Handle outliers
  - Group <blank>, “not entered”, “unknown”
  - Group low/high values together
  - Decide to eliminate values
- Dealing with data anomalies (in the Tennessee case, jurisdictional artifacts)
  - Metropolitan areas, not real jurisdictions, but medically significant
  - Single county wholly surrounded by a multi-county region
  - Extracting a large county from a multi-county region

Grouping Values



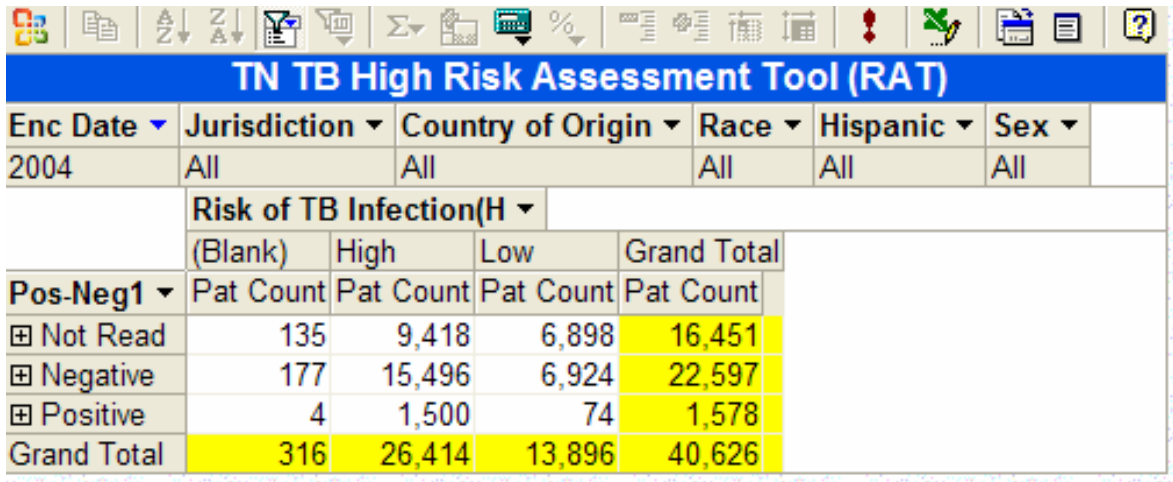
TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All	All

		Risk of TB Infection(H			
		(Blank)	High	Low	Grand Total
Pos-Neg1	Pos-Neg	Pat Count	Pat Count	Pat Count	Pat Count
⊟ Not Read	(Blank)	132	8,690	6,693	15,515
	Not Read	3	728	205	936
⊟ Negative	Negative	177	15,496	6,924	22,597
⊟ Positive	Positive	4	1,500	74	1,578
Grand Total		316	26,414	13,896	40,626

Grouped

Delete old detail

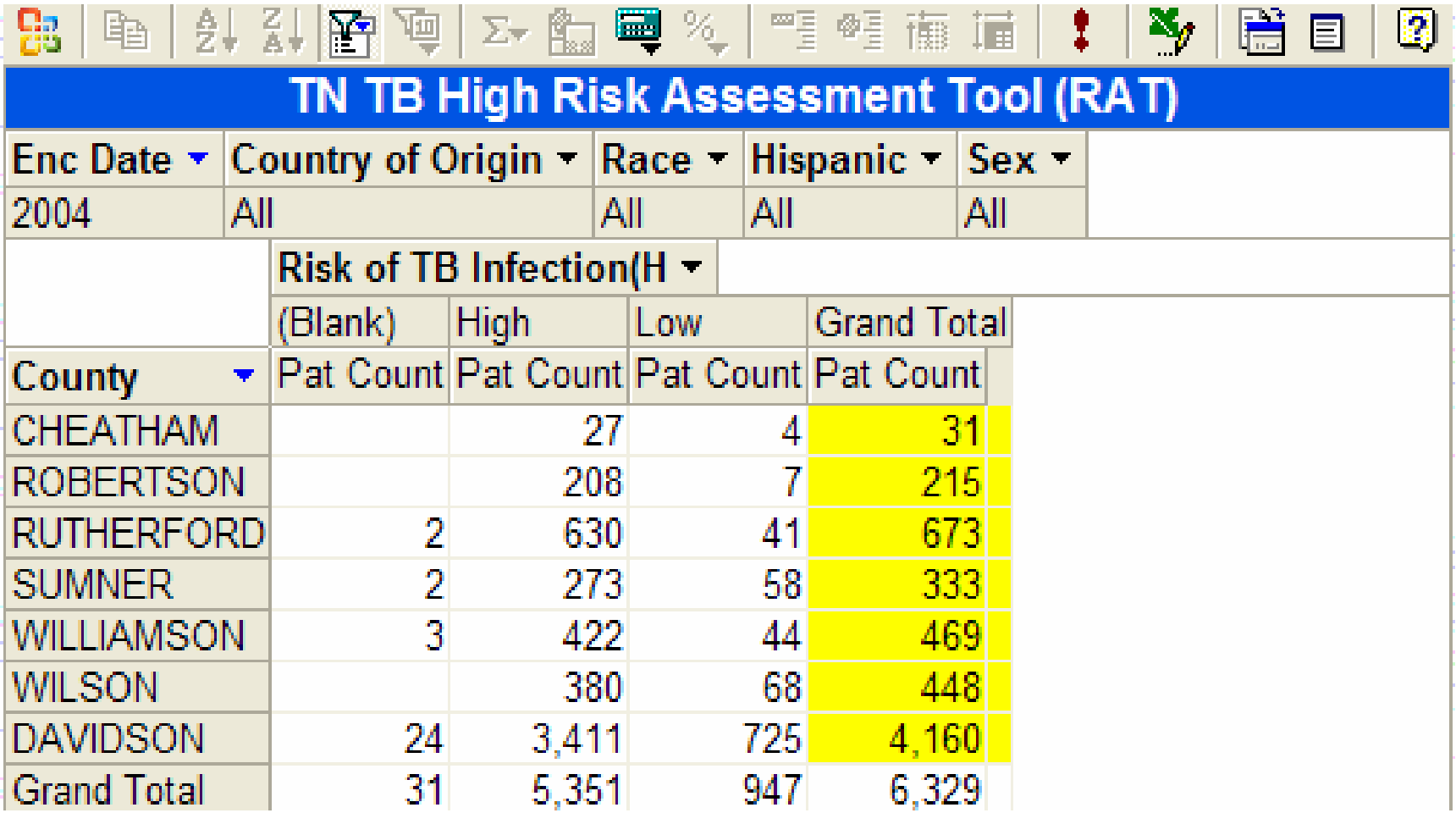


TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All	All

	Risk of TB Infection(H			
	(Blank)	High	Low	Grand Total
Pos-Neg1	Pat Count	Pat Count	Pat Count	Pat Count
⊞ Not Read	135	9,418	6,898	16,451
⊞ Negative	177	15,496	6,924	22,597
⊞ Positive	4	1,500	74	1,578
Grand Total	316	26,414	13,896	40,626

Regrouping Jurisdictions

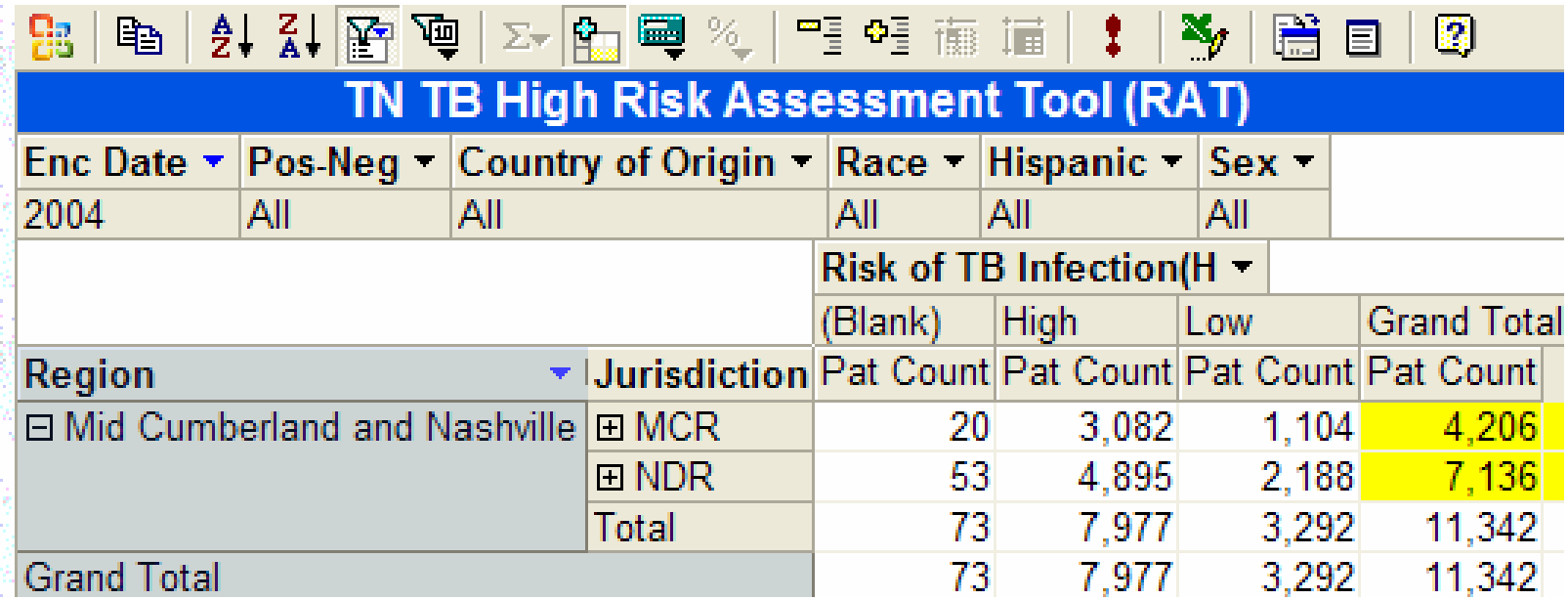


TN TB High Risk Assessment Tool (RAT)

Enc Date	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All

	Risk of TB Infection(H			
	(Blank)	High	Low	Grand Total
County	Pat Count	Pat Count	Pat Count	Pat Count
CHEATHAM		27	4	31
ROBERTSON		208	7	215
RUTHERFORD	2	630	41	673
SUMNER	2	273	58	333
WILLIAMSON	3	422	44	469
WILSON		380	68	448
DAVIDSON	24	3,411	725	4,160
Grand Total	31	5,351	947	6,329

Regrouping Jurisdictions



TN TB High Risk Assessment Tool (RAT)

Enc Date	Pos-Neg	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All	All

		Risk of TB Infection(H			
		(Blank)	High	Low	Grand Total
Region	Jurisdiction	Pat Count	Pat Count	Pat Count	Pat Count
⊟ Mid Cumberland and Nashville	⊞ MCR	20	3,082	1,104	4,206
	⊞ NDR	53	4,895	2,188	7,136
	Total	73	7,977	3,292	11,342
Grand Total		73	7,977	3,292	11,342

Questions ???

Contacts

Marion Kainer, M.D., M.P.H.

Marion.Kainer@state.tn.us

John A. Roberts

John.A.Roberts@state.tn.us

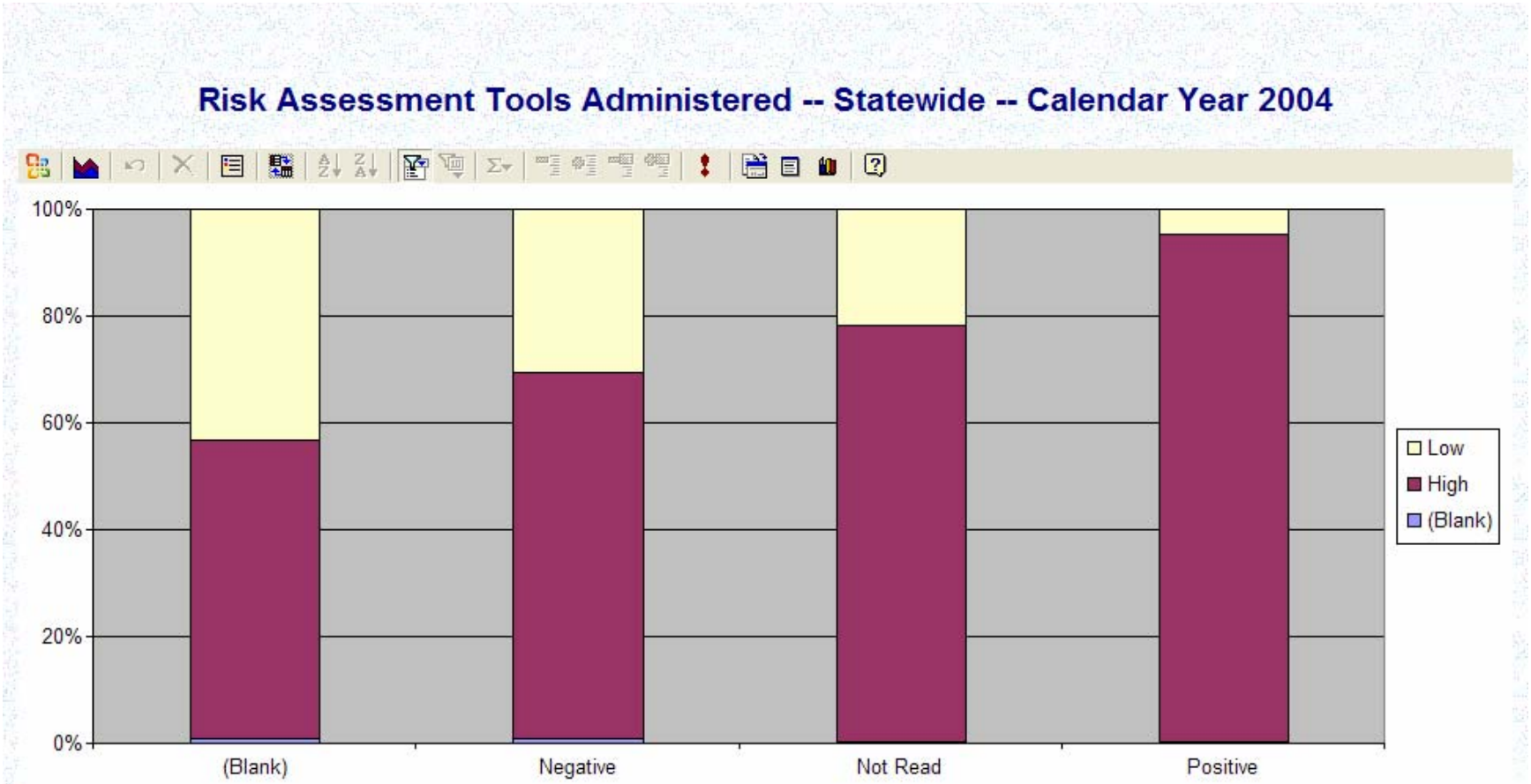
Gary Self

Gary.Self@state.tn.us

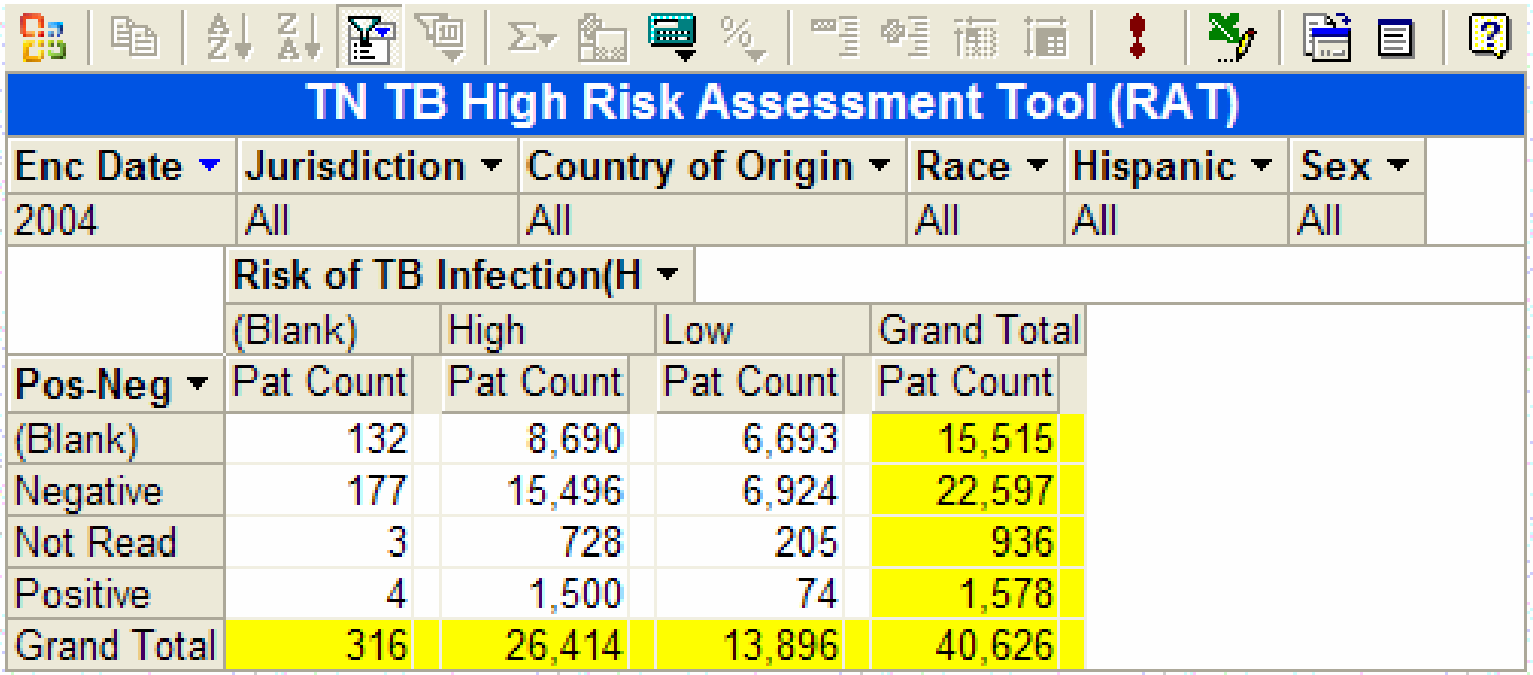
Additional Slides

- Cube nomenclature

The Chart


Risk Assessment Tools Administered -- Statewide -- Calendar Year 2004
100%
80%
60%
40%
20%
0%
(Blank)
Negative
Not Read
Positive
Low
High
(Blank)

The Table

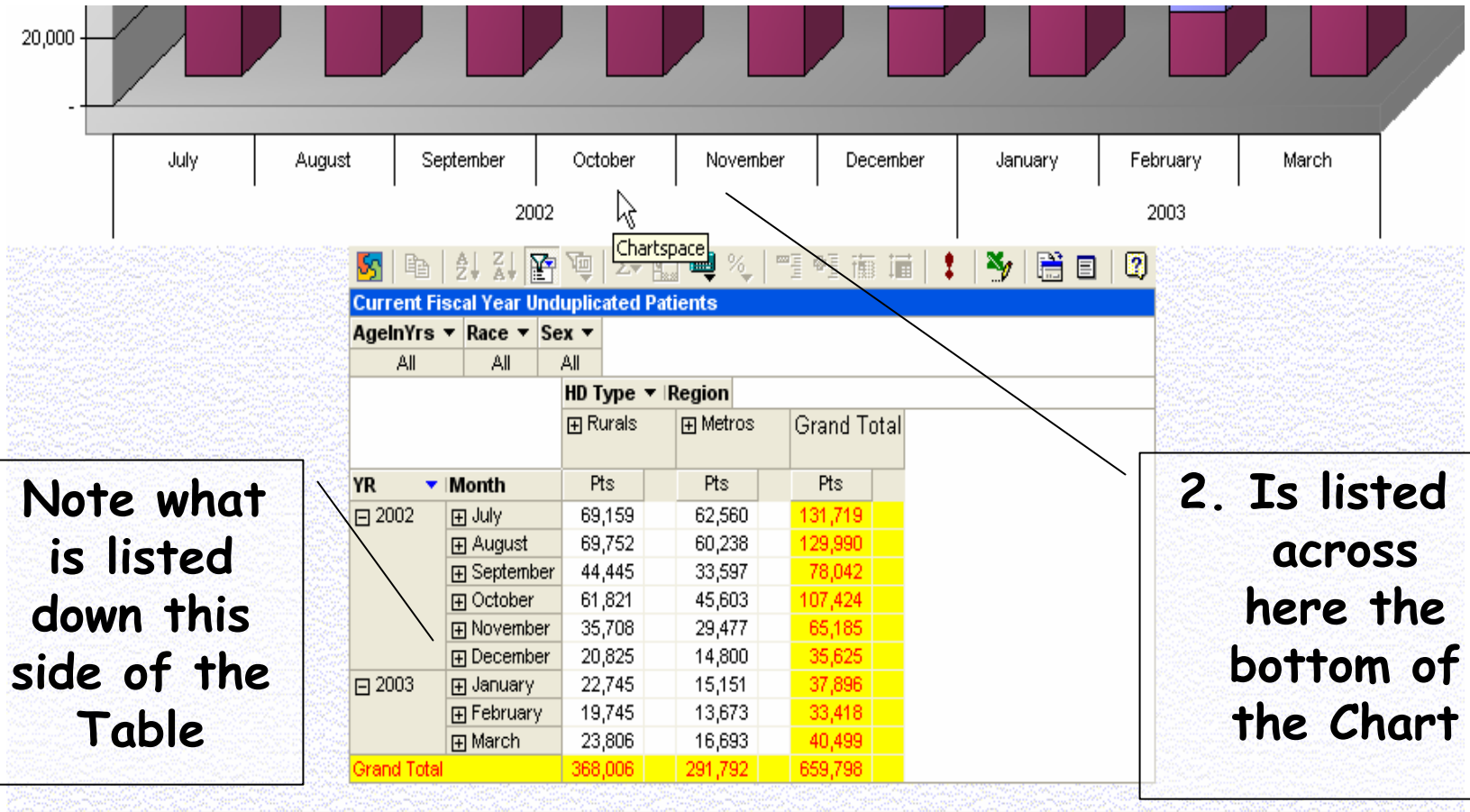


TN TB High Risk Assessment Tool (RAT)

Enc Date	Jurisdiction	Country of Origin	Race	Hispanic	Sex
2004	All	All	All	All	All

	Risk of TB Infection(H			
	(Blank)	High	Low	Grand Total
Pos-Neg	Pat Count	Pat Count	Pat Count	Pat Count
(Blank)	132	8,690	6,693	15,515
Negative	177	15,496	6,924	22,597
Not Read	3	728	205	936
Positive	4	1,500	74	1,578
Grand Total	316	26,414	13,896	40,626

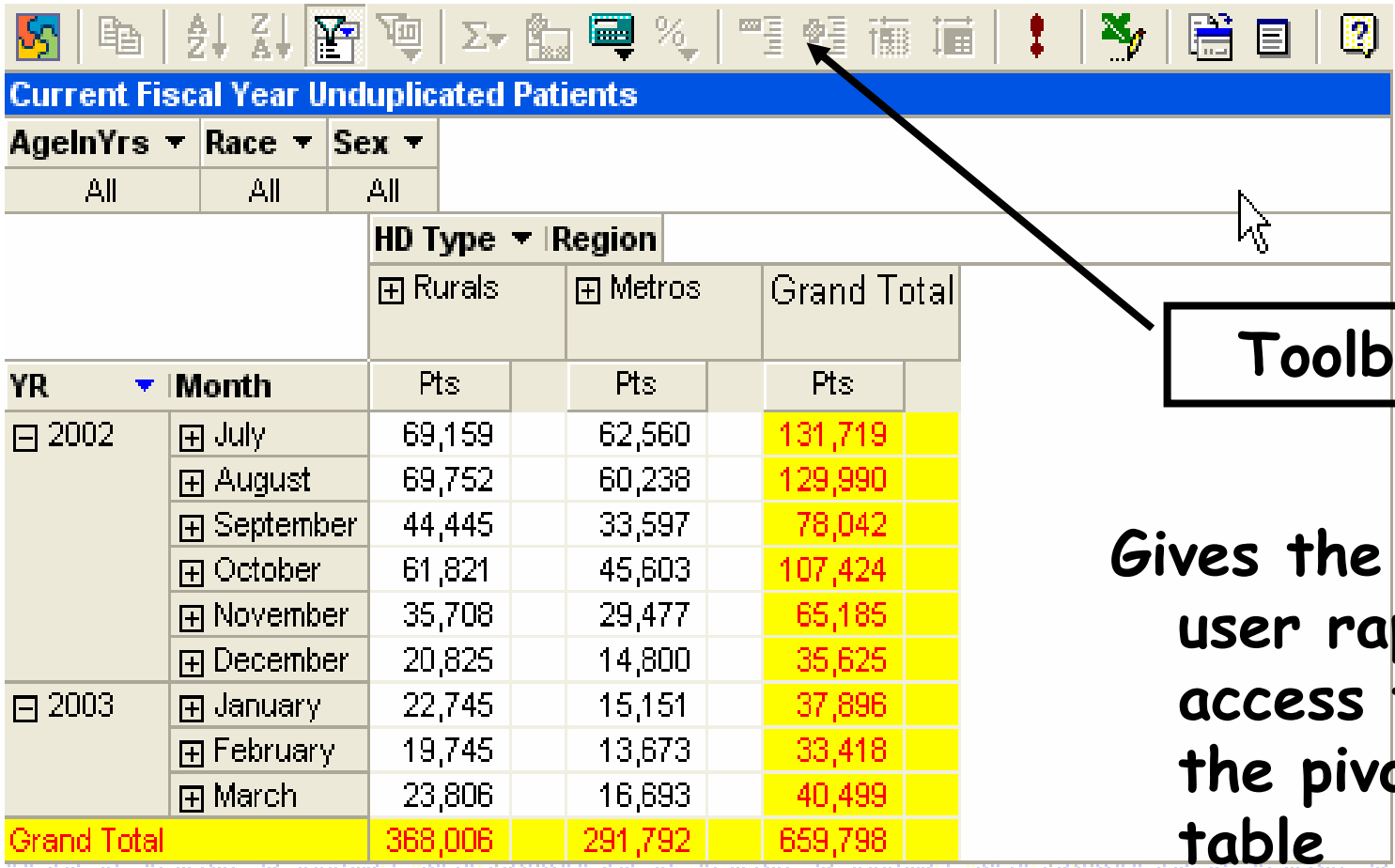
Getting Around



1. Note what is listed down this side of the Table

2. Is listed across here the bottom of the Chart

What is it?



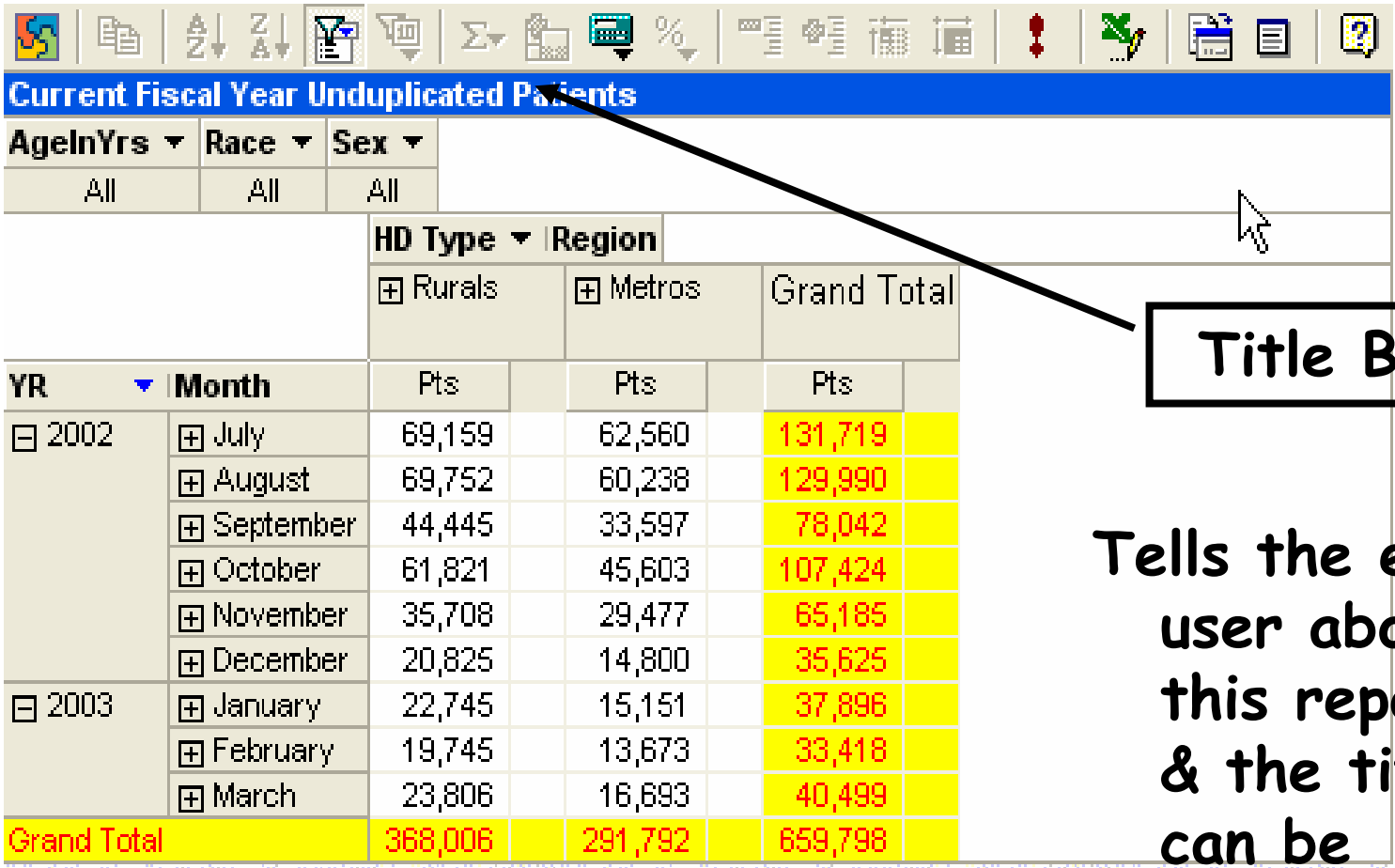
Current Fiscal Year Unduplicated Patients

AgeInYrs	Race	Sex
All	All	All

		HD Type: Rurals		HD Type: Metros		Grand Total	
YR	Month	Pts		Pts		Pts	
2002	July	69,159		62,560		131,719	
	August	69,752		60,238		129,990	
	September	44,445		33,597		78,042	
	October	61,821		45,603		107,424	
	November	35,708		29,477		65,185	
	December	20,825		14,800		35,625	
2003	January	22,745		15,151		37,896	
	February	19,745		13,673		33,418	
	March	23,806		16,693		40,499	
Grand Total		368,006		291,792		659,798	

Gives the user rapid access to the pivot table options

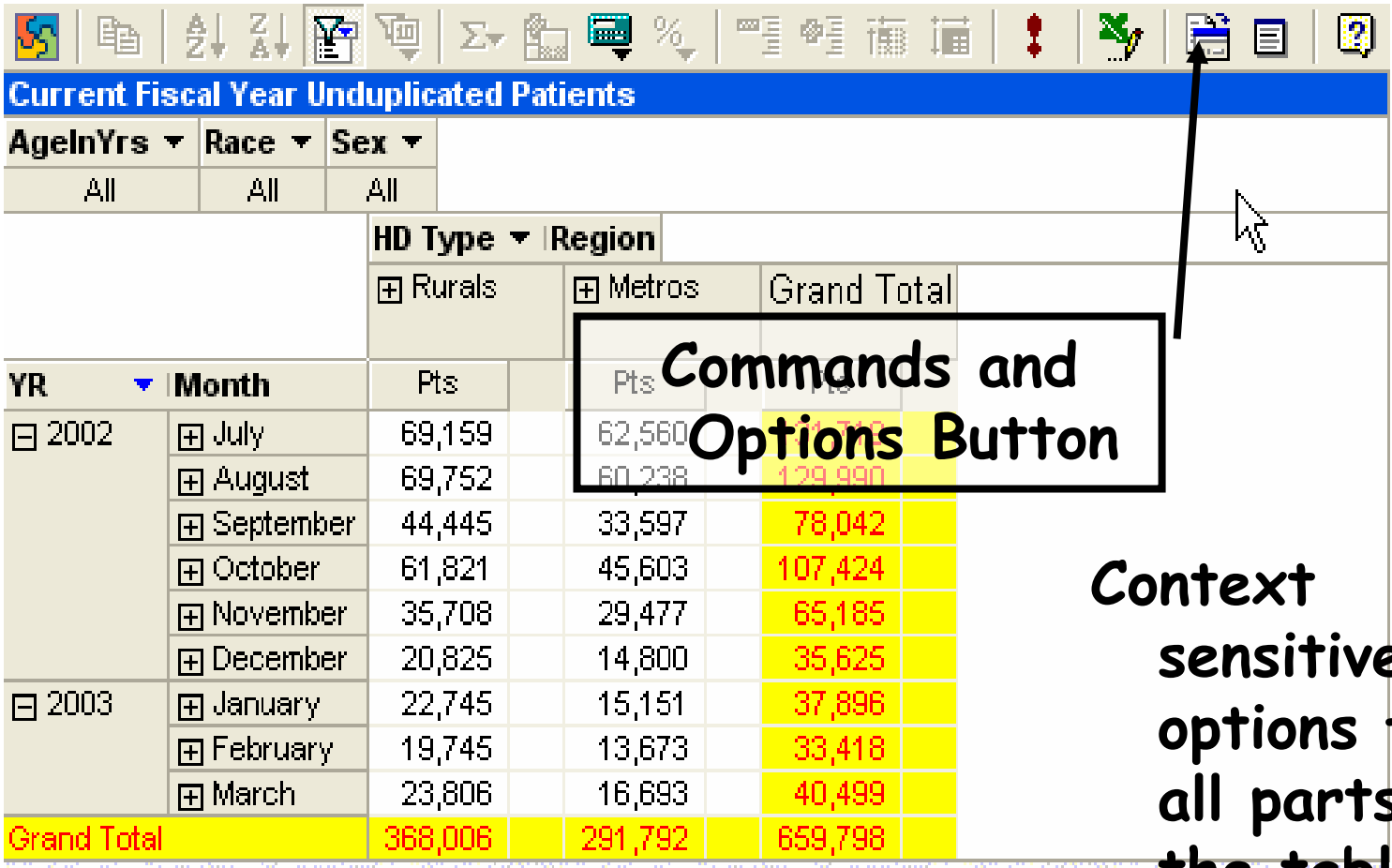
What is it?



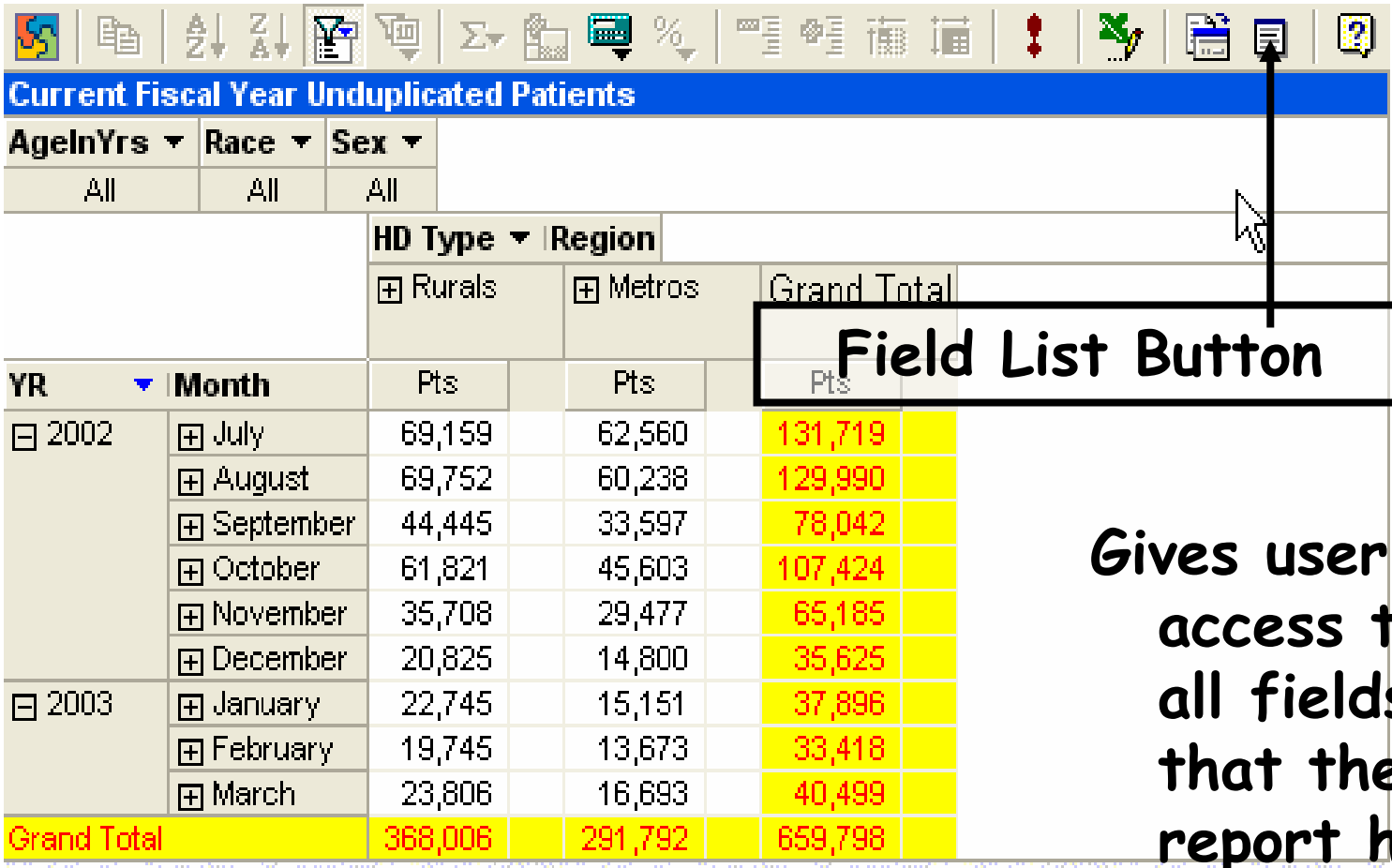
Title Bar

Tells the end user about this report & the title can be changed

What is it?

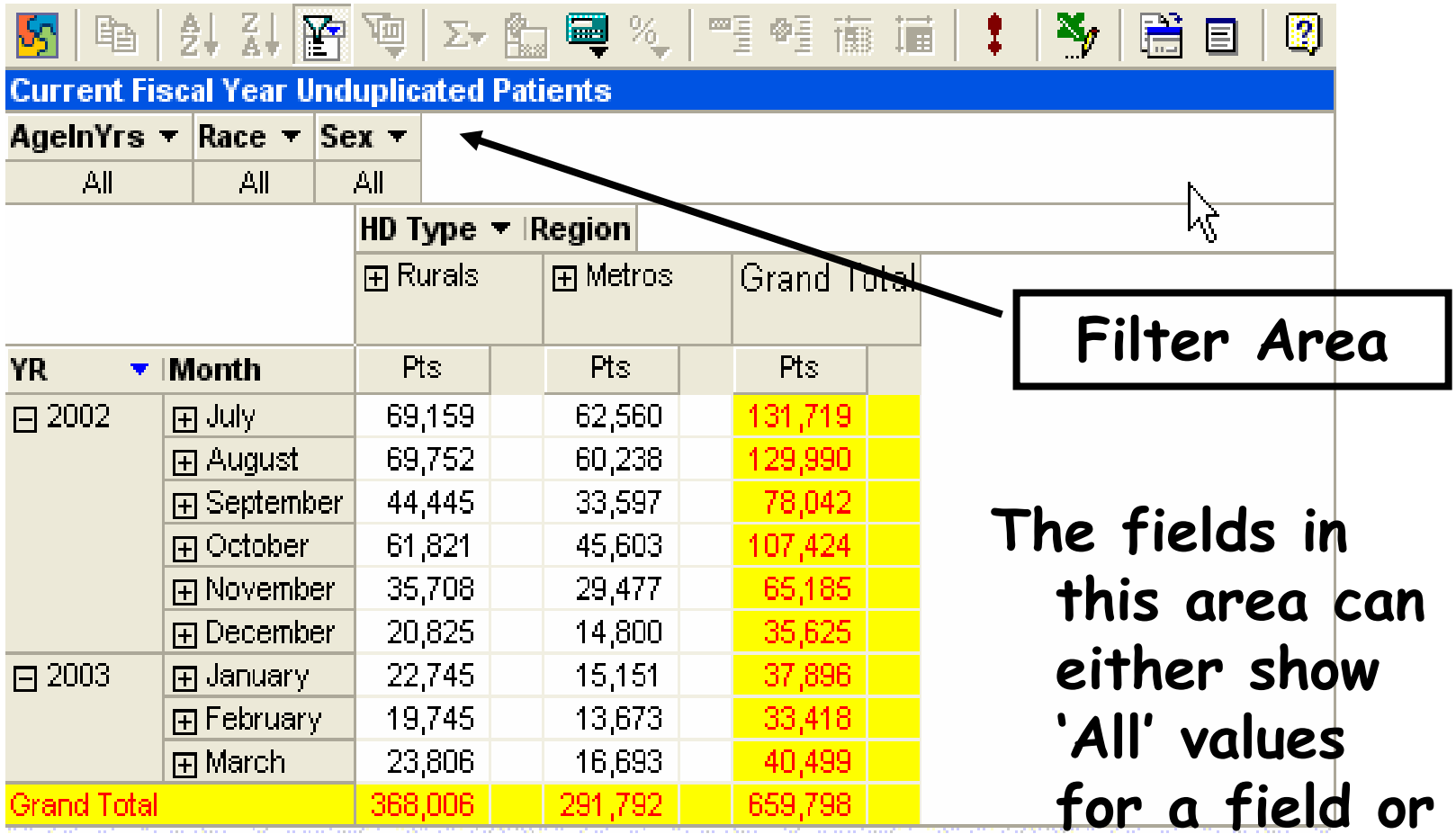

Current Fiscal Year Unduplicated Patients
AgeInYrs
Race
Sex
All
All
All
HD Type
Region
Rurals
Metros
Grand Total
YR
Month
Pts
Pts
2002
July
69,159
62,560
August
69,752
60,238
September
44,445
33,597
78,042
October
61,821
45,603
107,424
November
35,708
29,477
65,185
December
20,825
14,800
35,625
2003
January
22,745
15,151
37,896
February
19,745
13,673
33,418
March
23,806
16,693
40,499
Grand Total
368,006
291,792
659,798
Commands and Options Button
Context sensitive options for all parts of the table or chart

What is it?



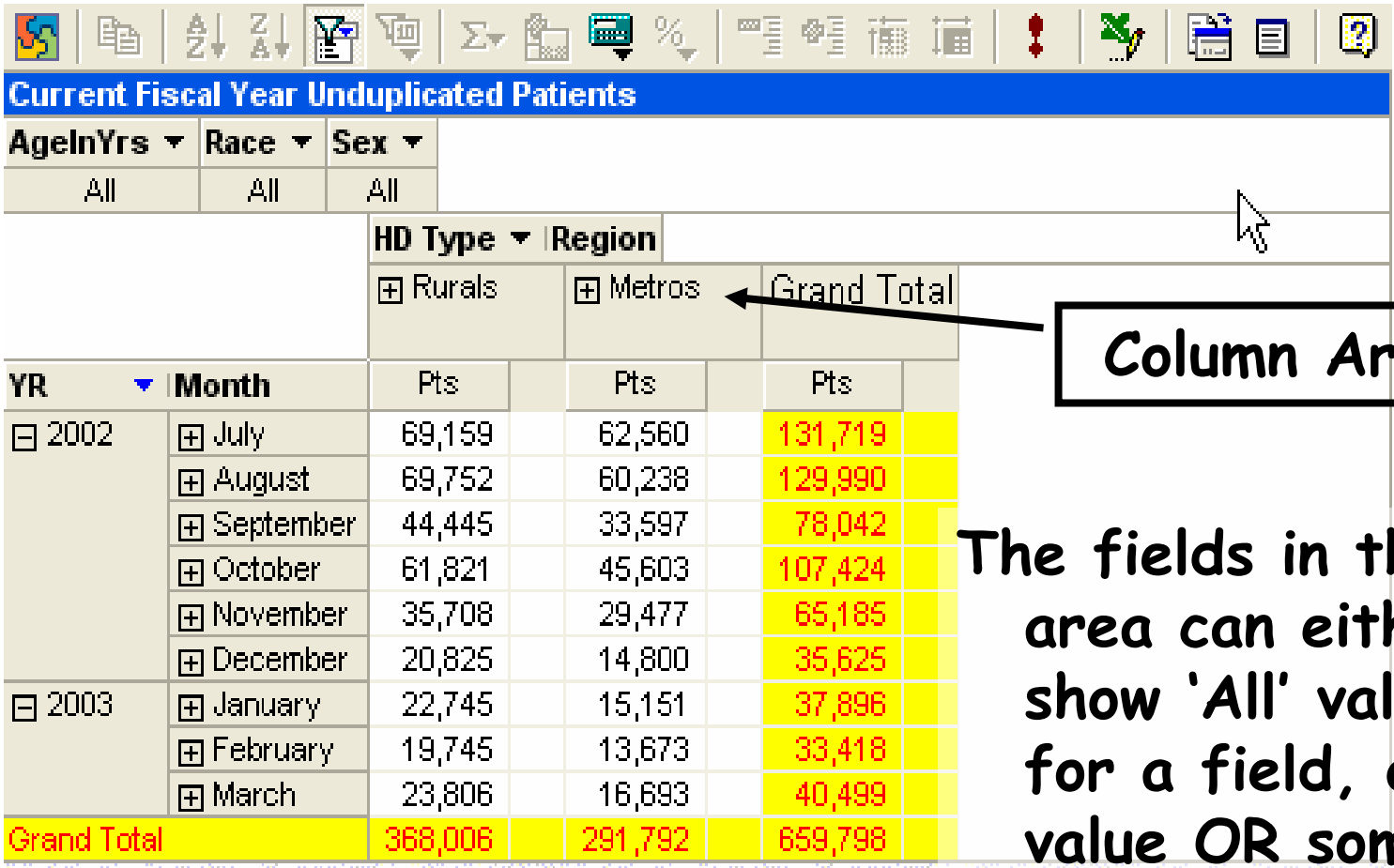
Gives user access to all fields that the report has available

What is it?



The fields in this area can either show 'All' values for a field or just one value for the field

What is it?



Current Fiscal Year Unduplicated Patients

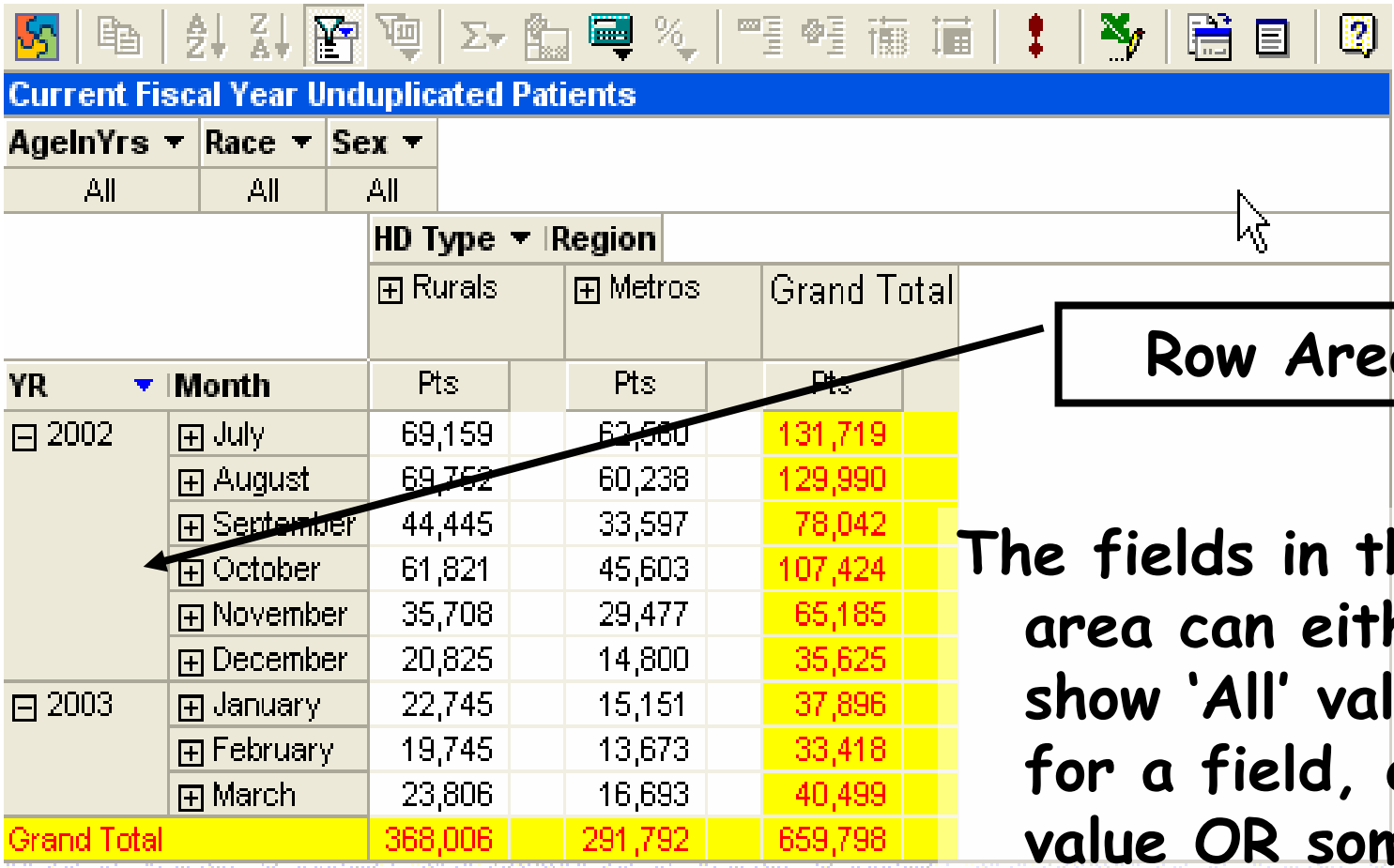
AgeInYrs	Race	Sex
All	All	All

	HD Type	Region			
	⊞ Rurals		⊞ Metros		Grand Total
YR	Month	Pts		Pts	Pts
⊟ 2002	⊞ July	69,159		62,560	131,719
	⊞ August	69,752		60,238	129,990
	⊞ September	44,445		33,597	78,042
	⊞ October	61,821		45,603	107,424
	⊞ November	35,708		29,477	65,185
	⊞ December	20,825		14,800	35,625
⊟ 2003	⊞ January	22,745		15,151	37,896
	⊞ February	19,745		13,673	33,418
	⊞ March	23,806		16,693	40,499
Grand Total		368,006		291,792	659,798

Column Area

The fields in this area can either show 'All' values for a field, one value OR some of the values for the field

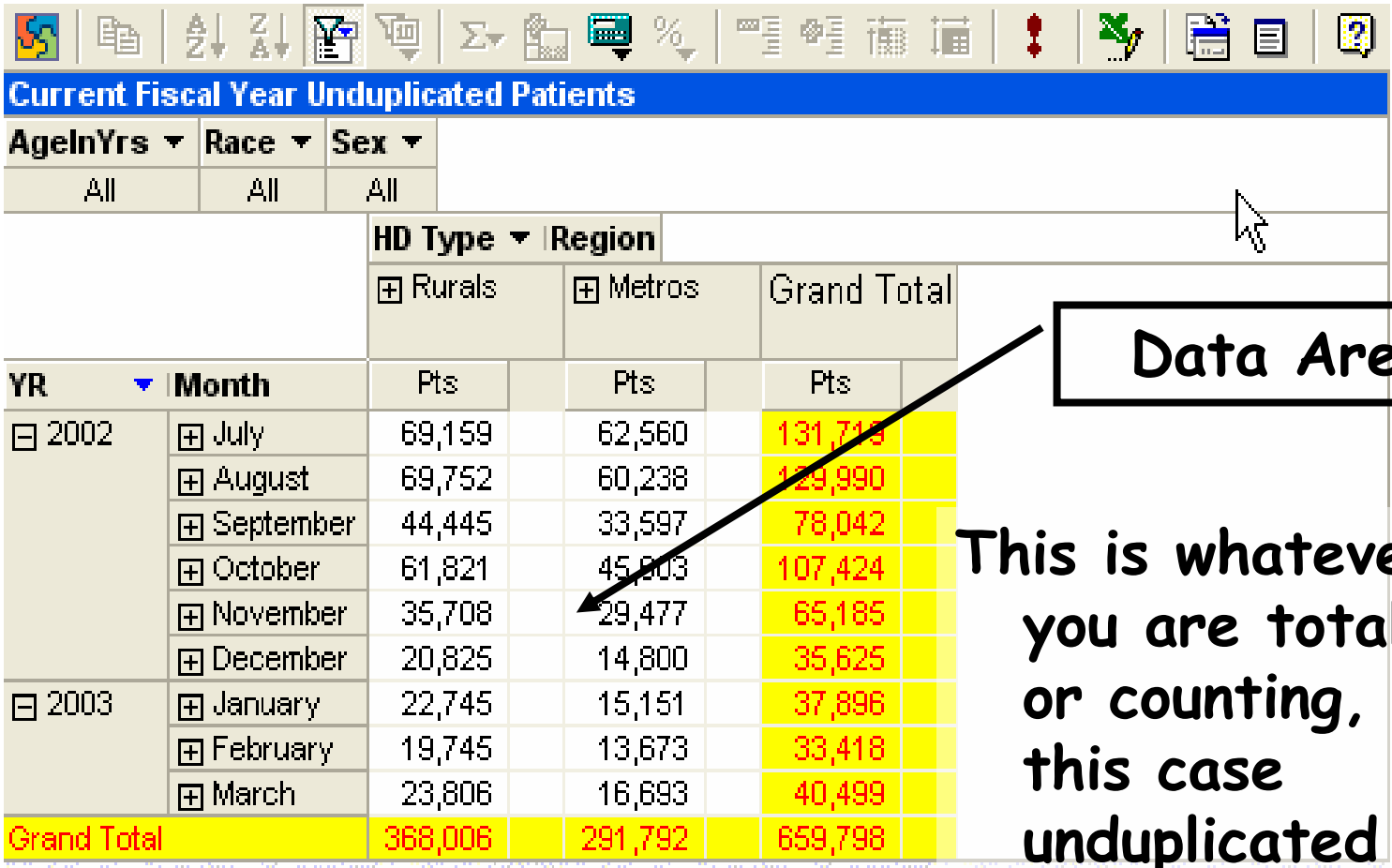
What is it?



Row Area

The fields in this area can either show 'All' values for a field, one value OR some of the values for the field

What is it?



Current Fiscal Year Unduplicated Patients

AgeInYrs	Race	Sex
All	All	All

		HD Type: Rurals	HD Type: Metros	Grand Total
YR	Month	Pts	Pts	Pts
2002	July	69,159	62,560	131,719
	August	69,752	60,238	129,990
	September	44,445	33,597	78,042
	October	61,821	45,603	107,424
	November	35,708	29,477	65,185
	December	20,825	14,800	35,625
2003	January	22,745	15,151	37,896
	February	19,745	13,673	33,418
	March	23,806	16,693	40,499
Grand Total		368,006	291,792	659,798

Data Area

This is whatever you are totaling or counting, in this case unduplicated patients